BioOne COMPLETE

Evolution of Weediness and Invasiveness: Charting the Course for Weed Genomics

Authors: Stewart, C. Neal, Tranel, Patrick J., Horvath, David P., Anderson, James V., Rieseberg, Loren H., et al.

Source: Weed Science, 57(5) : 451-462

Published By: Weed Science Society of America

URL: https://doi.org/10.1614/WS-09-011.1

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://bioone.org/journals/Weed-Science on 11 Oct 2024
Terms of Use: https://bioone.org/terms-of-use

Weed Science 2009 57:451–462

Evolution of Weediness and Invasiveness: Charting the Course for Weed Genomics

C. Neal Stewart, Jr., Patrick J. Tranel, David P. Horvath, James V. Anderson, Loren H. Rieseberg, James H. Westwood, Carol A. Mallory-Smith, Maria L. Zapiola, and Katrina M. Dlugosch*

The genetic basis of weedy and invasive traits and their evolution remain poorly understood, but genomic approaches offer tremendous promise for elucidating these important features of weed biology. However, the genomic tools and resources available for weed research are currently meager compared with those available for many crops. Because genomic methodologies are becoming increasingly accessible and less expensive, the time is ripe for weed scientists to incorporate these methods into their research programs. One example is next-generation sequencing technology, which has the advantage of enhancing the sequencing output from the transcriptome of a weedy plant at a reduced cost. Successful implementation of these approaches will require collaborative efforts that focus resources on common goals and bring together expertise in weed science, molecular biology, plant physiology, and bioinformatics. We outline how these large-scale genomic programs can aid both our understanding of the biology of weedy and invasive plants and our success at managing these species in agriculture. The judicious selection of species for developing weed genomics programs is needed, and we offer up choices, but no *Arabidopsis*-like model species exists in the world of weeds. We outline the roadmap for creating a powerful synergy of weed science and genomics, given well-placed effort and resources.

Key words: Bioinformatics, DNA sequencing, gene expression, genetic transformation, genomics, systems biology, weed biology.

Weedy and invasive species cause up to $100 billion in damage annually in crop and ecosystem function loss (Pimental et al. 2005), but the biological mechanisms responsible for their success remain poorly understood. Genomics is an approach to understanding biology that involves global analysis of gene organization, expression, and function at the whole-genome level (Hieter and Boguski 1997). Genomic tools offer unparalleled opportunities to dissect the genetic basis and evolution of traits associated with the success of weedy and invasive plants. Although many weed scientists already study complex features of biology that arise from gene activity across the genome, most researchers have not been able to make use of these genomic resources.

Genomics has language, assumptions, and conventions of its own, and these can pose a barrier to the uninitiated researcher. The broad scope of genomic research necessitates the use of high-throughput technologies, and these generate large data sets that are comprehensible only with the aid of computer analyses. This situation can be daunting to researchers who are not versed in the computational tools of BIOINFORMATICS (see Box 1), particularly because most genomics software developed to date is not user-friendly. Moreover, the high cost of incorporating genomics into research projects has historically prohibited the application of these technologies to nonmodel systems. To put the powerful tools of genomics to work for weed science, we must explore the nature of these information and technology gaps and how they might be closed.

Bridging the gap from genomics to weed science is not without precedent. A similar situation occurred with the emergence of molecular biology, which, in its early years, was viewed by many weed scientists as a separate and foreign discipline. Of course, today the techniques of molecular biology pervade all types of biological research and have provided tremendous insights into the biology of weeds, including their origin, dispersal, and mechanisms of control. In much the same way, genomic approaches promise to extend our insights again, this time beyond individual genes to the nature and evolution of complex traits and genomes as a whole (Yuan et al. 2008).

Although a number of reviews have been published on the use of genomics, molecular genetics, and biochemistry in weed science (Basu et al. 2004; Chao et al. 2005; Indergit et al. 2006; Stewart 2009; Yuan et al. 2007), the development of genomic tools and resources for weedy and invasive species lags far behind that for crops and model species. The question is how to focus the expertise and resources of the scientific community on achieving a set of common goals for weed genomics. A recent workshop was held to tackle the major issues related to developing a weed genomics research plan (Table 1) and to chart the course of a research agenda. The product of that workshop is provided here as a proposed roadmap for using modern research tools in weedy and invasive plant biology, especially to better understand the evolution of these traits. This article will review (1) key strategies for using genomic approaches to achieve the goals of weed science, (2) examples of successful research programs in this area, (3) candidate species for efficient leveraging of genomic resources, and (4) how weed scientists can move toward implementing this agenda in their research.

Genomic Approaches to Weed Science

Genomic technologies already have a proven record of advancing our understanding of basic animal and plant biology. For example, ecological genomics has tackled issues of organismal response to environment, genetic variation, and adaptation (Karrenberg and Widmer 2008; Thomas and Klaper 2004; Wu et al. 2008). The challenge for the weed

DOI: 10.1614/WS-09-011.1

* First author: Department of Plant Sciences, University of Tennessee, Knoxville, TN 37996; second author: Department of Crop Sciences, University of Illinois, Urbana, IL 61801; third and fourth authors: U.S. Department of Agriculture–Agricultural Research Service, Fargo, ND 58105; fifth and ninth authors: Botany Department, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; sixth author: Department of Plant Pathology, Physiology and Weed Science, Virginia Tech, Blacksburg, VA 24061; seventh and eighth authors: Department of Crop and Soil Science, Oregon State University, Corvallis, OR 97331. Corresponding author's E-mail: nealstewart@utk.edu

Downloaded From: https://bioone.org/journals/Weed-Science on 11 Oct 2024
Terms of Use: https://bioone.org/terms-of-use

Table 1. Questions that should be addressed to develop a strategic and comprehensive weed genomics research plan.

1. Which weed species, and how many, should we choose to develop as models for genomics research?
2. What are the key tools and resources that should be developed for the model weeds?
3. How do we best use comparative genomic approaches to learn what genomic attributes make plants weedy?
4. How do we bridge the gaps between disciplines of weed science, genomics, and evolutionary biology?
5. How do we ensure weed genomics research is translational, such that the results lead to improved weed management?
6. How do we obtain funding to support comprehensive weed genomics research?

science community is how to maximize the use of genomic approaches to answer questions that are important to weed biology. Genomics must help us understand the traits that have made weeds successful colonizers and troublesome pests, as well as how these features evolve, because we know that weeds can adapt quickly (Barrett 1983; VanGessel 2001). There are two main genomic approaches to understanding the genetic basis of the enhanced performance of weedy and invasive species: genome analyses and trait analyses. These two approaches complement one another, with genome analyses generating insights into loci and traits of interest, their evolutionary context, and interactions with other loci, and trait-based analyses providing insights into the nature and function of the genes underlying focal weedy traits (Figure 1). These approaches share an interrelated set of genomic tools (Figure 2; Box 1).

Genome Analyses: Population Genomics. Analyses of the genome itself include both population genomics and functional genomics (Figure 1). Population genomics refers to the assessment of genetic variation and differentiation within loci across the genome (Stinchcombe and Hoekstra 2008). This requires gathering genomic data from multiple individuals. Although WHOLE-GENOME SEQUENCING of model organisms will be essential for providing frameworks for assembly and annotation of related individuals, population genomics must make use of a host of different methods for determining genome-wide patterns of sequence variation. These methods range from indirect assays with MOLECULAR MARKERS (Kane and Rieseberg 2007; Neale and Ingvarsson 2008; Wood et al. 2008) to direct sequencing of EXPRESSED SEQUENCE TAGS (ESTs; see also, GENE-SPACE SEQUENCING) using NEXT-GENERATION SEQUENCING technologies (Mardis 2008). Patterns of genetic similarity among weedy and related

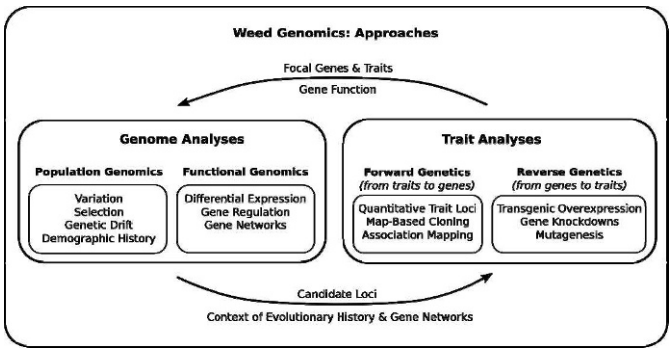

Figure 1. Major approaches to weed genomics, the insights gained from each analysis, and how different strategies complement one another.

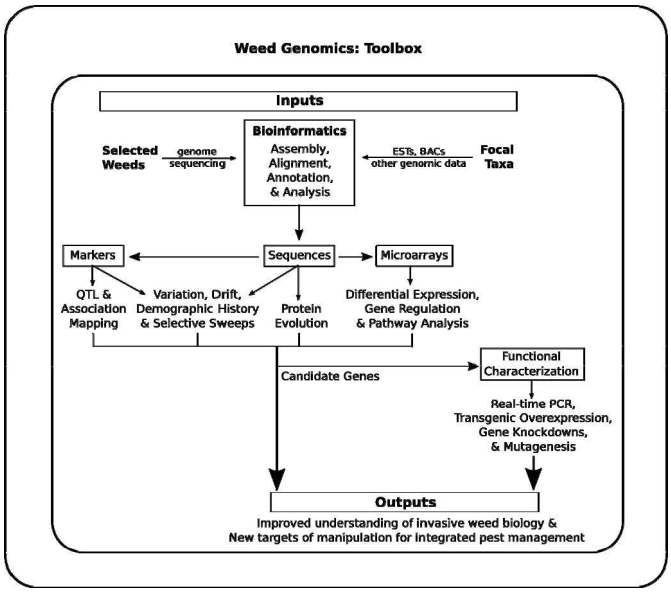

Figure 2. A toolbox for weed genomics, illustrating the techniques currently available (see Box 1) and how certain tools contribute data that enable or inform other tools.

populations can then be used to reveal important aspects of population history, including the origin of weedy and invasive populations, their history of expansion, their propensity for gene flow, and their tendency to hybridize with related taxa (Kane and Rieseberg 2007; Zayed and Whitfield 2008). Genome-wide scans are particularly sensitive methods for detecting gene flow and hybridization, critical issues in weed science, where the acquisition of locally adapted traits or resistance to chemical and biological controls might be rare in occurrence but high in importance (Dlugosch and Whitton 2008; Ellstrand and Scierenbeck 2000; Whitney et al. 2006).

An especially powerful application of genomics is in the identification of targets of selection, including both artificial selection, imposed by control efforts, and natural selection, for colonizing ability and adaptation to local environments. Detection of putative selection from molecular-marker scans relies on outlier analyses, in which loci that show the greatest reduction in diversity (a selective sweep), or greatest genetic distance (diversifying selection), or both, are viewed as possible targets of selection. However, marker-based scans appear to have a high false-positive rate (Wiehe et al. 2007), so these studies are best viewed as providing a ranked list of candidate loci. Also, although marker-based approaches offer an inexpensive means of identifying candidate loci, they fail to detect the actual sites targeted by selection (although see Wood et al. 2008).

A broader and more powerful array of methods is available for detecting signs of selection in sequence data (Wright and Gaut 2005). These include methods of testing for selective sweeps via reduced variability (Hudson-Kreitman-Aguadé [HKA] test; Hudson et al. 1987) or mutation frequency distribution shifts (Tajima's *D* test; Tajima 1989) and testing for protein evolution via increased nonsynonymous substitution rates (nonsynonymous [Ka] to synonymous [Ks] ratio test, Yang 1998; McDonald-Kreitman test, McDonald and Kreitman 1991). Although these methods are less prone to false positives than marker-based approaches, again, it is probably best to employ them for providing ranked lists of

Downloaded From: https://bioone.org/journals/Weed-Science on 11 Oct 2024
Terms of Use: https://bioone.org/terms-of-use

candidate genes. It is the observation of the same genetic changes in invasive populations that have independent origins that will provide the strongest evidence for identifying specific genes or mutations as being responsible for weedy and invasive traits. Parallel evolution of functional groups of genes might also reveal consistent trade-offs that contribute to invasion success, even if particular evolutionary pathways differ among populations or species (as observed in weedy sunflowers, Kane and Rieseberg 2008; Lai et al. 2008). Once loci under selection have been identified in species that are polymorphic for weedy and invasive behaviors, changes in these genes can be analyzed at a broader phylogenetic scale to better understand why weeds are concentrated in some groups of plants but not in other seemingly similar taxa.

Genome Analyses: Functional Genomics. Functional genomics includes the study of genome-wide patterns of gene expression (Hieter and Boguski 1997). It is possible to make quantitative comparisons of genomic expression patterns across species and populations by printing complimentary DNAs (cDNAs) or oligonucleotides onto MICROARRAYS and probing them with the transcriptomes (messenger RNA [mRNA] extractions) of different plants. Microarrays can be made for model species and used to survey expression in related species because heterologous microarray-hybridization experiments have been successful in species with divergence times as great as 65 million years (Renn et al. 2004; Taji et al. 2004). However, expression data are most easily interpreted if nucleic acid hybridizations are conducted using a microarray developed from the same species. Gene expression and sequence data can be obtained simultaneously using next-generation sequencing technologies. The comparisons made with these techniques can identify loci that are differentially expressed by weedy genotypes, suggest trade-offs in physiological responses to different environments or control measures, and reveal correlated responses among networks of interacting loci (Yuan et al. 2008).

Detecting the selection for weediness genes from expression data is more challenging than from sequence data. If weedy and nonweedy populations are exchanging genes, then significant expression differences (measured in a uniform environment) are probably a consequence of selection, although maternal environments, particularly temperature differences, could also affect gene expression (Blödner et al. 2007; Johnsen et al. 2005). As with sequence data, the strongest evidence of selection comes from parallel expression shifts in weedy populations that have independent origins (Lai et al. 2008). A major issue in the interpretation of expression data is whether a significant expression change is a direct target of selection or a side-product of selection on other genes (pleiotropy). In principle, it should be feasible to distinguish between these alternatives by determining the regulatory basis of the expression changes: *cis*-regulated changes are more likely to be the direct product of selection, whereas *trans*-regulated changes are more likely to result from pleiotropy (Landry et al. 2007).

Expression data are well-suited to identifying physiological trade-offs experienced by weeds as they invade different environments or face various control measures. Microarray experiments have already provided valuable insight into physiological processes related to weediness (Horvath and Clay 2007; Horvath et al. 2003, 2006a, 2008). By surveying genomic expression in plants grown under various conditions, we can understand which loci or classes of genes are up-regulated or down-regulated in different environments. The latest generation of arrays even allows identification of individual genes within larger gene families, permitting assessments of how divergent functions of gene paralogs might contribute to the broad ecological tolerances of many weedy species (Chao et al. 2005; Kim et al. 2008).

Using genomic expression data in combination with genome sequences, as previously described, also provide opportunities for detecting short transcription-factor binding sites shared between clusters of coordinately regulated genes (Tatematsu et al. 2005). These clusters could play important roles in regulating various weedy traits, and such characterization is an important complement to trait-based analyses (see below). Importantly, characterization of transcription factors could also provide molecular targets for novel herbicide development.

Trait Analyses. Trait analyses, the second major genomic approach (Figure 1), focus on the genetics of traits that are hypothesized a priori to contribute to weediness or invasiveness, such as competitiveness, high fecundity, delayed germination, the ability to reproduce vegetatively, and herbicide tolerance or resistance (Gressel 2002). Trait-based analyses include both forward genetics, which starts with the phenotype and moves toward gene identification, and reverse genetics, which starts with a gene and moves toward identifying the phenotype it affects. These approaches would be greatly facilitated by the creation of model weed systems, ranging from full genome sequencing to the development of permanent mapping populations to transgenics (Figure 2). The fact that 80.6% of the genes in *Arabidopsis* are also found in rice (*Oryza sativa* L.; Yu et al. 2002) underscores the potential for many of the genes and physiological processes controlling weedy and invasive traits to be shared among model and nonmodel species. However, a 20% (or even 5%) difference is sizable. There is a need to develop the genomics of species that are diverse with respect to life history and phylogeny.

Whole-genome sequencing and the development of population genomic markers can be used to perform forward genetics, including mapping of QUANTITATIVE TRAIT LOCI (QTL) in controlled crosses and ASSOCIATION MAPPING of loci to phenotypes in natural populations. These techniques are critical for identifying the genetic basis of key traits, such as herbicide resistance and plant parasitism. By understanding their genetic basis, we will be able to track the evolution of these traits, as well as their occurrence, inheritance, and dispersal. This information, in turn, provides the ability to predict responses to different control measures, to genetically tailor management to weeds, and to modify crops genetically for resistance to parasites. Focal genes for trait analyses will also prompt further genome-level analyses to understand the evolutionary context and interactions of these key genes.

The connection between particular loci and phenotypes cannot be confirmed without reverse genetics, where the effects of genes are demonstrated directly by GENETIC TRANSFORMATION. Manipulation of genes in plants can be done by transgenic overexpression, gene knockdown analysis, or mutagenesis. For example, a putative herbicide-resistance gene cloned from a resistant genotype might be overexpressed in an otherwise susceptible genotype, followed by subsequent

Downloaded From: https://bioone.org/journals/Weed-Science on 11 Oct 2024
Terms of Use: https://bioone.org/terms-of-use

herbicide challenge. Or that same gene's expression could be knocked down in the resistant genotype, challenged with herbicide, and tested for conversion to herbicide susceptibility. The combination of transformation and susceptible- and resistant-biotypes would be valuable for screening putative, nontarget, herbicide-resistance targets from other species in overexpression assays. In fact, genomic approaches have already been used in the search for herbicide target sites by high-throughput knockout of genes (Lein et al. 2004). Efficient transformation systems will be a necessary component of any genomic analysis because the biological significance of identified putative weediness genes must be verified by investigating their effect in the species of interest.

Benefits to Weed Management

Ultimately, the practical goal of weed genomics is to aid in weed management. Support for genomic research is dependent upon its application to the needs of end users and its benefits to agriculture and the environment. Genomics and related molecular techniques have the potential to provide these practical benefits by increasing our ability to identify traits that contribute to weediness, to find new effective and environmentally sound control measures, and to predict evolutionary responses to our management practices (Anderson 2008).

Historically, it has been difficult to precisely define the traits and genes that make a species particularly weedy and invasive. Invasiveness in a particular environment depends on the genomic constitution of the weed species and on the environment at the site of introduction. For example, an agronomic weed might have succeeded by accumulating domestication traits, such as the loss of dormancy and shattering that mimic a crop (Warwick and Stewart 2005). In contrast, for invasive weeds of wild or natural areas, success may be based on the retention of those traits (Lai et al. 2006). Understanding the sources of genetic variation for these traits and their rapid adaptation to different environments could lead to the ability to predict whether and where a weed will become invasive (Prentis et al. 2008). Genome scans that compare gene-sequence diversity across populations can be used to reveal which loci are associated with success in different environments and to determine the sources of variation in those traits (i.e., standing variation vs. new mutations).

Herbicide resistance is undoubtedly the most important trait affecting long-term control of weedy populations. Genomics provides powerful opportunities to elucidate the action of herbicides (Eckes et al. 2004), the evolution of herbicide resistance, and the occurrence, inheritance, and dispersal of herbicide-resistance genes. Extensive information, mainly using DNA sequencing and single nucleotide polymorphism (SNP) analysis, has already been used in research examining the molecular mechanisms of target-site herbicide resistance (Devine and Shukla 2000; Tranel and Wright 2002). However, fewer nontarget-site resistance mechanisms have been elucidated at the molecular level because of the more complicated basis of this type of resistance and the limited genomic information available for weedy species (Yuan et al. 2007). Global gene-expression profiling techniques, such as microarrays, are a powerful tool for studying the molecular responses to herbicide application (Lee and Tranel 2008; Raghavan et al. 2005, 2006) and can be especially valuable in identifying nontarget herbicide-resistance mechanisms (Yuan et al. 2007). Molecular markers have been used to investigate single vs. multiple origins of herbicide resistance, gene flow, and the frequency of resistant alleles in weed populations, which are factors that strongly influence weed management strategies (Bodo Slotta 2008). Genomics approaches might be able to finally provide a mechanistic understanding of the utility of herbicide rotations vs. herbicide mixtures for prevention of resistance and of the effect of low doses or high doses on the evolution of resistance. A mechanistic understanding would allow us to predict when and where a particular practice (e.g., low dose vs. high dose) would be correct (Gardner et al. 1998). The identification of pathways involved in herbicide response may also suggest novel molecular targets for herbicide development.

Parasitic weeds are among the most difficult weeds to manage because of the physical and physiological interactions of these species with their host plants. Genomic techniques can aid in identifying host genes that naturally provide resistance to parasitic weeds or genetic pathways critical for parasitic infection. Metabolomics and proteomics could be used to identify the unique features of plants that are naturally resistant to parasites, which could lead to identification of the genes responsible for resistance (Gressel 2008). Additionally, such studies are likely to suggest the pathways and genes in the parasite that are required for infection, offering targets for new control measures.

We know that these weediness traits can evolve in response to control measures, and genomics can help us to identify sources of variation in weedy populations and to predict their evolutionary responses to control. Genome-scale surveys of molecular markers can quantify gene flow and the frequency of hybridization, which can affect weed management practices (Bodo Slotta 2008; Tranel and Wright 2002). In particular, gene flow and hybridization have recently become popular areas of study because of the movement of herbicide-resistance genes both from naturally evolved resistance genes and from transgenes. However, the effects of gene flow on traits, such as salt- or drought-tolerance that could increase a weed's fitness, have not, as yet, to our knowledge, been addressed (Mallory-Smith and Zapiola 2008).

Finally, coupling estimates of gene flow with rates of adaptation in weediness traits would be particularly powerful for guiding management. For example, comparing selective pressures on genes in weeds sampled from different cropping systems would provide information on the roles that agricultural fields, fallow fields, and natural areas play in the maintenance of heritable adaptive traits. This information aids in the design of weed management systems: A high migration rate with a low adaptation rate would require different management than if both migration and adaptation rates were high. In the latter case, it would be important to change management practices more quickly to minimize opportunities for the weed to adapt.

Successful Examples of Weed Genomics Research

Evolutionary Population Genomics in the Compositae Family. As far as we are aware, evolutionary population genomic methods have thus far only been applied to weeds in the sunflower (Compositae syn. Asteraceae) family (Stevens 2007). Evolutionary genomic studies have been feasible in this group because of the development of EST libraries and micro-

Downloaded From: https://bioone.org/journals/Weed-Science on 11 Oct 2024
Terms of Use: https://bioone.org/terms-of-use

arrays for several weeds in the family (Barker et al. 2008; Broz et al. 2007; Church et al. 2007; Lai et al. 2006). Most of this work has been done through the Compositae Genome Project (http://compgenomics.ucdavis.edu/compositae_index.php), which has been funded by the now defunct U.S. Department of Agriculture (USDA) Initiative for Future Agriculture and Food Systems (IFAFS) program and more recently by the National Science Foundation (NSF) Plant Genome Program, with the goal of developing genomic tools and resources for this large and economically important family.

Three studies from the Compositae Genome Project illustrate both the promise and challenges of evolutionary population genomics. A scan of 106 microsatellite (simple sequence repeat [SSR]) loci for evidence of selection in wild and weedy sunflower (*Helianthus annuus* L.) populations identified several loci that have swept through one or more weedy populations. The scans employed SSRs located within ESTs, which have the advantage of providing candidate genes that are known to be expressed and tightly linked to each locus. Although most of the putative sweeps appear to represent examples of local or regional adaptation, rather than selection for weediness per se, one gene (a heat shock protein) exhibited independent sweeps across weed populations from across the United States and, thus, appears to represent a "weedy gene" (Kane and Rieseberg 2008). Likewise, microarray experiments using a first-generation cDNA array (3,100 unique genes; Lai et al. 2006) identified 165 genes, representing about 5% of total genes on the array, which showed differential expression in one or more weed populations (Lai et al. 2008). Two functional categories of genes were significantly overrepresented: response to stress and response to biotic or abiotic stimulus. However, the most intriguing finding was that genes with consistent expression differences across all four weed populations assayed were mostly down-regulated, implying trade-offs with other functions and potential adaptation to more benign conditions.

More recently, the Roche GS-FLX (454) next-generation sequencing platform[1] has been employed to sequence normalized cDNAs from 10 native and 10 invasive yellow star-thistle (*Centaurea solstitialis* L.) genotypes (K. Dlugosch, M. Barker, Z. Lai, and L. Rieseberg, unpublished data). An average of 89,000 200-bp reads and 32,000 unigenes were obtained per genotype or about 71 Mbp/plate. Despite fairly low redundancy, preliminary assemblies and analyses indicate that approximately 2,000 unigenes can be scanned for evidence of selection. As in sunflower, genes involved in stress responses predominate among those showing evidence of selection, a result consistent with a trade-off hypotheses for weed evolution, which posits that plants are unable to be highly stress tolerant and highly competitive (or reproductive) simultaneously (Grime 1977). Should this observation prove to be general, it would provide one of the first mechanistic explanations for the evolution of weediness. Evolutionary genomics, within and among species, thus offers powerful new opportunities to identify common mechanisms facilitating the success of weedy and invasive plants, with important implications for the management of these species.

Comparative and Functional Genomics in Leafy Spurge. Leafy spurge (*Euphorbia esula* L.) is a member of the Euphorbiaceae family that contains important agronomic crops, such as cassava (*Manihot esculenta* Crantz), castorbean (*Ricinus communis* L.), and rubber tree [*Hevea brasiliensis* (Willd. ex A. Juss.) Müll. Arg.], as well as horticultural species, such as poinsettia (*Poinsettia pulcherrima* Willd. ex Klotzsch). Leafy spurge has been considered as a model to study seed and adventitious root bud dormancy in perennial dicot weeds (Chao et al. 2005). However, early attempts to garner support and funding to initiate a genomic program for leafy spurge met with little success. To overcome some of the financial hurdles, potential collaborators working on related species were identified. Several research groups realized that a significant understanding of the conservation and diversity of genes between members of Euphorbiaceae was lacking but saw the potential for identifying economically important genes and physiological/developmental processes common to multiple members of this plant family; this initiated the pursuit of a coordinated large-scale effort for developing genomics resources in multiple Euphorbiaceae species, including cassava, rubber tree, and leafy spurge.

Preliminary collaborations demonstrated good cross-species utility of genomic resources (Anderson et al. 2004). Thus, based on a common goal of generating a Euphorbiaceae-specific microarray, various in-house resources, collaborative agreements, and small, competitive grants were used to develop a low-cost program that resulted in the production of about 23,000 unique leafy spurge sequences (Anderson et al. 2007) and about 9,000 unique cassava sequences (Lokko et al. 2007). These ESTs have been annotated and organized for the construction of Euphorbiaceae-specific DNA microarrays, which represent in excess of 23,000 unigene set, including 19,015 leafy spurge unigenes and 4,129 unigenes from cassava. The development and use of these high-density microarrays are enhancing our understanding of genes and genetic networks associated with traits that make perennial weeds, such as leafy spurge, so invasive and difficult to control (Horvath et al. 2008). The success of these initial collaborations have resulted in further success stories, which include (1) grants, through the U.S. Department of Energy–Joint Genome Institute (DOE-JGI), to sequence the genome of cassava; (2) development of two sets of 96 SSR markers from cassava ESTs that are being used in breeding programs in Africa (interestingly, 80% of these SSRs work in amplifying leafy spurge DNA); and (3) construction of cassava-specific oligo arrays through Agilent Technologies.[2] It is still too early to know the full agricultural benefit from the original collaborative initiative, but many research groups are currently using these valuable genomic tools. It is evident that pooling resources and developing collaborative projects are essential for developing programs in weed genomics.

Weed Candidates for Genomics

The question of which or how many candidate weeds should be chosen for answering the fundamental and practical questions of interest to weed scientists is critical for developing a roadmap for genomic exploration of weed biology and ecology. A single weed that lends itself to genomic manipulation and that can be used to answer most of the questions of interest to weed scientists would be ideal for focusing funding and intellectual efforts, a strategy proven by the mouseear cress [*Arabidopsis thaliana* (L.) Heynh.] plant model. However, no single species can encompass all weedy

Downloaded From: https://bioone.org/journals/Weed-Science on 11 Oct 2024
Terms of Use: https://bioone.org/terms-of-use

traits. Instead of posing one or two model weeds that might be prescribed to the community of researchers and stakeholders, we will discuss the characteristics needed for weed candidates. It is clear that even if such a single-weed model existed, it is unlikely that mechanisms imparting weediness to any single species will have analogous mechanisms in all other weeds, given the diversity in weedy species. Thus, it is inevitable that more than one candidate weed is needed for developing a robust weed genomics program. In choosing candidate species for large-scale genomics research, several factors are important (Basu et al. 2004; Chao et al. 2005):

(1) Candidate weeds considered for development of a weed genomics program must have a foundation of previous research, providing critical preliminary data and demonstrating their feasibility as study systems. Recent research efforts by the weed science and invasive plant community provide an indicator of weeds amenable to further study (Figure 3).
(2) Candidate weeds must pose a large threat. Weeds that infest a broad range of habitats and that are troublesome over geopolitical boundaries are likely to inspire support for funding from multiple governmental and nongovernmental agencies. Fortunately for the quest to identify a limited number of candidate species, the world's most troublesome and well-studied weeds (Holm et al. 1997) also display a large number of classic weedy characteristics, with most exhibiting more than 70% of the 14 weediness traits described by Baker (1974).
(3) Candidate weeds must be amenable to genome-scale studies. Ideally, model species should have small genomes (See Figure 4) or genomes with significant synteny to sequenced genomes, permitting detailed comparative studies and inferences across study systems (Basu et al. 2004). Advances in genomic technologies have made it feasible to study plants with complex genetics, such as wheat (*Triticum aestivum* L.), but model species with small genomes will remain the most tractable and affordable systems for concerted research efforts.
(4) Candidate species should be easily manipulated through genetic transformation. Genetic transformation is a very useful tool for elucidating the links between genotype and phenotype (Figures 1 and 2) and has already proven useful in advancing weed science (Halfhill et al. 2007).

As stated above, any given model is unlikely to encompass all weedy traits in a manner perfect for genomic analysis. Numerous potential model species have previously been suggested based on the above or similar criteria (Basu et al. 2004, Chao et al. 2005). More recently, nearly 100 national and international weed scientists interested in exploring genomics of weeds were tasked with making a short list of candidates (WSSA 2008). Among the weeds considered were pigweed (*Amaranthus* L. ssp.), Johnsongrass [*Sorghum halepense* (L.) Pers.], leafy spurge, jointed goatgrass (*Aegilops cylindrica* Host), purple (*Cyperus rotundus* L.) and yellow nutsedge (*Cyperus esculentus* L.), common ragweed (*Ambrosia artemisiifolia* L.), nightshades (*Solanum* L. ssp.), and many others. The group came to consensus that a diverse suite of species are worth pursuing further. What follows are some examples.

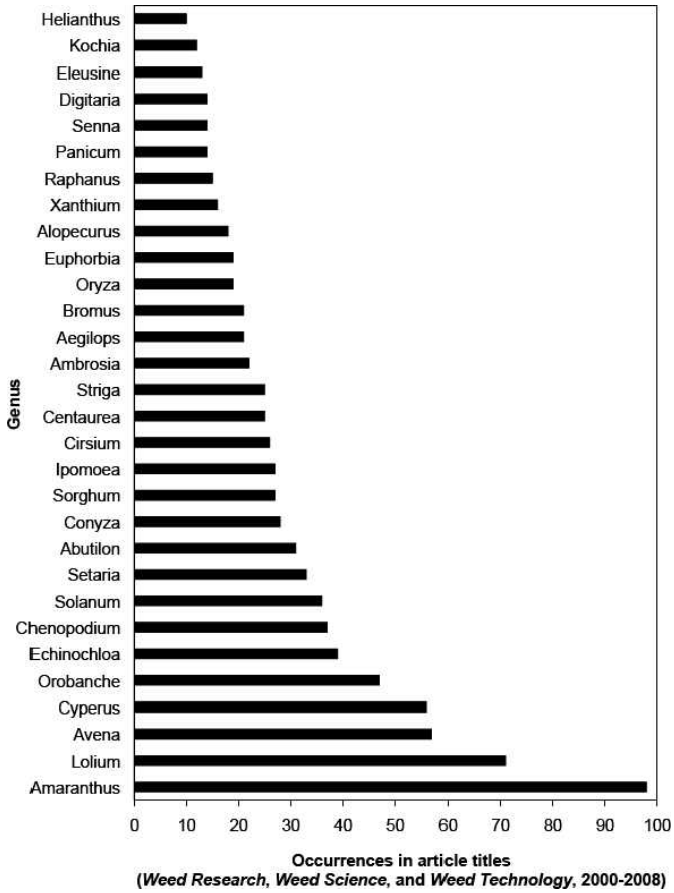

Figure 3. The most significant weed-containing genera as determined by their occurrences in weed science literature. Sixty genera expected to have high occurrences were chosen from the Weed Science Society of America's Composite List of Weeds (http://www.wssa.net/Weeds/ID/WeedNames/namesearch.php) and used to search article titles published in *Weed Research*, *Weed Science*, or *Weed Technology* from 2000 through 2008. Hits to article titles containing only crop species with the genus name were not counted. Only those genera having 10 or more occurrences are shown. Data amended from Tranel and Trucco (2009).

Ryegrass. Ryegrass (*Lolium* L. spp.; Poaceae) is among the best-studied weed genera (Figure 3) An extensive EST database (Sawbridge et al. 2003) and cDNA microarrays (Ciannamea et al. 2006) already exist for these weeds. Ryegrasses have widespread distributions (Charmet et al. 1996) and are problematic in numerous habitats, including agricultural, range, and recreational settings (Bossard et al. 2000). Ryegrasses have numerous weedy characteristics, such as varying levels of seed dormancy, high fecundity, ability to propagate by seed and tillers, potential for cross-species hybridization, and herbicide resistance (Basu et al. 2004). *Lolium* ssp. have genome sizes of about 4,067 Mbp (Figure 4; Evans et al. 1972) making them poor candidates for full genome sequencing, but synteny with, and genomic resources developed for other Poacea, make gene-space sequencing a possibility. Also, ryegrasses are generally small enough to grow and study in a limited laboratory facility. Finally, transformation systems have been developed for ryegrass, although the transformation is inefficient (Wu et al. 2005).

Canada Thistle. Canada thistle [*Cirsium arvense* (L.) Scop.; Compositae] is generally dioecious; however, some true

Downloaded From: https://bioone.org/journals/Weed-Science on 11 Oct 2024
Terms of Use: https://bioone.org/terms-of-use

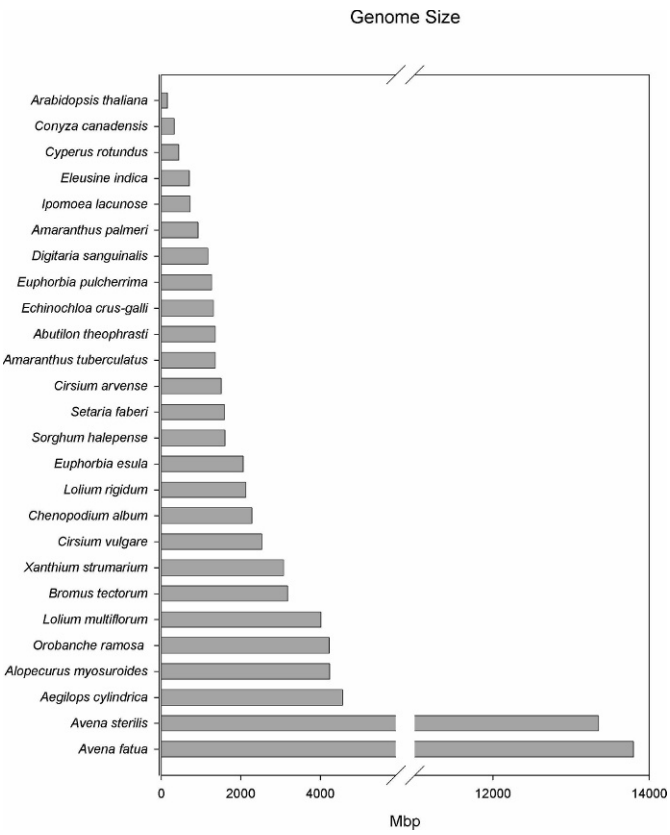

Figure 4. Genome size (Mpb/1C) estimates of various weedy species and *Arabidopsis thaliana*. Most estimates are from the Kew Gardens C-value database (http://data.kew.org/cvalues/), updated in 2005. Estimates for weeds not found in the Kew database were obtained from recent publications: *Euphorbia esula* (Chao et al. 2005) and *Amaranthus* species (Rayburn et al. 2005). The *Orobanche ramosa* estimate was published in Weiss-Schneeweiss (2005). The *Conyza canadensis* estimate was obtained from flow cytometry (Peng, Yuan, Tranel, and Stewart, unpublished data).

hermaphroditic plants have been observed (Heimann and Cussans 1996 and references therein), and thus it presents an excellent model to study the effect of mating-system variation on invasiveness (Barrett et al. 2008). This species also reproduces vegetatively, offering another mode of reproduction for comparison and the opportunity to propagate clonal experimental plants. Various genomic resources, including extensive EST collections and microarrays, have been developed for related species in the Compositae (Barker et al. 2008) and should enable efficient development of genomic-based tools for Canada thistle. Canada thistle is a diploid with a genome size of about 1,519 Mbp (Figure 4; Bennett and Leitch 2003) Again, its large genome precludes it from being rapidly sequenced. Several members of the Compositae have been transformed (Malone-Schoneberg et al. 1994; Michelmore et al. 1987; Narumi et al. 2005; among others), but some species, such as sunflower, are very recalcitrant against transformation. Thistle (*Cirsium* Mill. spp.) transformation is unknown.

Canadian Horseweed. Canadian horseweed [*Conyza canadensis* (L.) Cronquist; Compositae] is a nuisance weed that is highly selfing. It was the first dicot weed known to evolve glyphosate resistance and has the most widespread distribution of glyphosate-resistant biotypes of all weeds. Although the genus has not been the focus of intensive research (Figure 3), it is becoming more of an agricultural concern because of its rapid and widespread resistance evolution. It is the most attractive weed for whole-genome sequencing because it has the smallest genome of all surveyed weeds: 335 Mbp (Figure 4; Peng, Yuan, Tranel, and Stewart, unpublished data). It is very amenable to genetic transformation (Halfhill et al. 2007) and would be amenable to reverse-genetics approaches. Its transcriptome has recently been sequenced using GS-FLX (454) technology, which produced 411,962 raw reads, averaging 233 bp, yielding a total data size of 95.8 Mb (Peng, Yuan, Tranel, and Stewart, unpublished data).

Pigweeds. Pigweeds (*Amaranthus* spp.; Amaranthaceae) are the most cited (Figure 3) and, arguably, the most troublesome weed pests in many agricultural settings. In addition, they are rapidly evolving herbicide resistance, in some cases, becoming resistant to multiple herbicides in the same plant (Patzoldt et al. 2005) Most pigweeds are monoecious (bisexual individuals with unisexual flowers), but some species are dioecious. *Agrobacterium*-mediated transformation has been performed in Prince-of-Wales feather (*Amaranthus hypochondriacus* L.; Jofre-Garfias et al. 1997), but the most important weed species have not been transformed The genome sizes of *Amaranthus* spp. weeds are moderately sized, ranging from approximately 900 Mbp for Palmer amaranth (*Amaranthus palmeri* S. Wats.) to 1,400 Mbp for tall waterhemp [*Amaranthus tuberculatus* (Moq.) Sauer; Figure 4; Rayburn et al. 2005). Waterhemp genomic resources have expanded rapidly in the past 6 mo. A GS-FLX (454) genomic DNA run produced 160,000 sequencing reads with an average read length of about 270 nucleotides, yielding a total of about 43 Mbp (P. J. Tranel, unpublished data). A 454-transcriptome run yielded 483,225 raw reads, with an average length of 232 bp, and a total data size was 114.8 Mbp (P. J. Tranel, unpublished data).

Implementation

As described above, successful inroads into weed genomics are already being made, and they clearly demonstrate that implementation of such programs will require the leveraging of resources from related species, especially crops, as well as extensive collaborations among researchers. Work on related species can provide useful genomic tools directly (Horvath and Clay 2007; Horvath et al. 2006b), as well as biological mechanisms that might apply across taxa (e.g., mechanisms regulating perennial dormancy in model plant species could be extended to the study of dormancy regulation in weeds). Importantly, both the leveraging of existing agricultural model species and the direct funding of weed genomics will require significant consumer and stakeholder input and support at the political level. Thus, it will be critical to raise awareness about the benefits of genomics for weed science.

A key factor that will influence the perceived benefits of weed genomics is economic: the value of the information gained compared with the cost of the work. Pooling resources from crop, weed, and other agricultural communities maximizes return on investment. Weed scientists have much to offer in the form of the compelling biological problems posed by weeds. Weeds encompass a wide range of biological traits that are both scientifically interesting and economically

Downloaded From: https://bioone.org/journals/Weed-Science on 11 Oct 2024
Terms of Use: https://bioone.org/terms-of-use

important. Moreover, although the study of traits such as herbicide resistance has obvious profitability within weed science alone, work on broader genomic analyses and other traits such as dormancy and allelopathy may be easier to justify for weed science if they illuminate the biology of other species of value. Reductions in sequencing costs are also improving the economics of weed genomics. Most sequencing conducted to date has used the relatively expensive Sanger technology (dideoxynucleotide sequencing),[3] such that genome sequencing of a plant species required massive financial inputs. As next-generation sequencing (Box 1) becomes routine and increasingly efficient, it will be easier to bridge the technology gap between genomics and weed science.

Collaborations among groups with different research foci will capitalize on a broad array of resources, generate the large body of research needed to establish model species, and allow weed scientists to build teams that can synergize different types of expertise. We can expect molecular weed science laboratories to find synergies to that make rapid progress toward well-defined goals.

Conclusions

Small inroads are already being made in weed genomics; however, a critical mass of scientists is needed to fully realize its potential. Weed scientists must initiate collaborations with genomics-oriented researchers and bioinformaticians, bringing together disparate areas of expertise and leveraging a broader array of financial resources for these large projects. It is also imperative that funding agencies recognize that the use of genomic approaches focused on major weed species, although not a panacea, will offer novel solutions and provide tangible benefits to science. For basic biology, the study of weed genomics offers a window into the world of rapid plant evolution and stress physiology. For weed management, nontarget-site herbicide resistance, in particular, mechanisms of parasitism and allelopathy, and evolution of invasiveness are weed science issues ripe for genomic-level analyses. Fundamental knowledge of the genetic underpinnings of what makes a plant a weed will provide for new management strategies to mitigate the negative effects of weedy and invasive plants on food production and habitat destruction.

No single species will encompass all of the myriad weediness traits, nor could it serve to answer all of the weed management questions. Thus, we argue that weed genomics should not be limited to a single species. In fact, the highest return on investment could be realized with initiating parallel projects that span the range of weediness traits and plant genera that are important to agriculture and the environment. Subsequent comparative genomics approaches could lead to strong inferences of important weed mechanisms, such as herbicide resistance and dormancy.

Possibly the most significant mile marker on the road map to weed genomics involves the next generation of scientists: We must ensure that graduate students and postdoctoral researchers are instilled with enthusiasm for weed science and receive the breadth of training needed to conduct their own genomic analyses. Therefore, the onus is on current weed scientists to develop collaborative weed-genomics research projects that will serve as training grounds and on funding agencies to provide the needed financial support. Only after scientists are both comprehensively trained in genomics and knowledgeable of weed management issues will we fully get the immense benefits weed genomics has to offer.

Sources of Materials

[1] GS-FLX (454) next-generation sequencing platform, Roche, Grenzacherstrasse 124, CH-4070, Basel, Switzerland.

[2] Cassava oligo arrays, Agilent Technologies, Inc., 5301 Stevens Creek Blvd., Santa Clara, CA 95051.

[3] Sanger dideoxynucleotide sequencing technology, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, U.K.

Acknowledgments

We are grateful for funding from the USDA-NRI to hold the workshop on the evolution of weedy and invasive plant genomics as part of the 2008 annual meeting of the Weed Science Society of America and also to the WSSA for also providing financial support for this workshop. We express our appreciation to members of our laboratories who not only generated the data that make this new line of investigating possible, but who also have enabled us to have the time and inspiration to write such a article. Thanks to Mat Halter for rendering Figure 4.

Literature Cited

Anderson, J. V. 2008. Emerging technologies: an opportunity for weed biology research. Weed Sci. 56:281–282.

Anderson, J. V., M. Delseny, and M. A. Fregene, et al. 2004. An EST resource for cassava and other species of Euphorbiaceae. Plant Mol. Biol. 56:527–539.

Anderson, J. V., D. P. Horvath, and W. S. Chao, et al. 2007. Characterization of an EST database for the perennial weed leafy spurge: an important resource for weed biology research. Weed Sci. 55:193–203.

Baker, H. G. 1974. The evolution of weeds. Annu. Rev. Ecol. Syst. 5:1–24.

Barker, M. S., N. C. Kane, A. Kozik, R. W. Michelmore, M. Matvienko, S. J. Knapp, and L. H. Rieseberg. 2008. Multiple paleopolyploidizations during the evolution of the Asteraceae reveal parallel patterns of duplicate gene retention after millions of years. Mol. Biol. Evol. 25:2445–2455.

Barbazuk, W. B., J. A. Bedell, and P. D. Rabinowicz. 2005. Reduced representation sequencing: a success in maize and a promise for other plant genomes. Bioessays 27:839–848.

Barrett, S.C.H., R. I. Colautti, and C. G. Eckert. 2008. Plant reproductive systems and evolution during biological invasion. Mol. Ecol. 17:373–383.

Barrett, S.C.H. 1983. Crop mimicry in weeds. Econ. Bot. 37:255–282.

Basu, C., M. D. Halfhill, T. C. Mueller, and C. N. Stewart, Jr. 2004. Weed genomics: new tools to understand weed biology. Trends Plant Sci. 9:391–398.

Bennett, M. D. and I. J. Leitch. 2003. Plant DNA C-values database. www.kew.org/cval.

Blödner, C., C. Goebel, I. Feussner, C. Gatz, and A. Polle. 2007. Warm and cold parental reproductive environments affect seed properties, fitness, and cold responsiveness in *Arabidopsis thaliana* progenies. Plant Cell Environ. 30:165–175.

Bodo Slotta, T. A. 2008. What we know about weeds: insights from genetic markers. Weed Sci. 56:322–326.

Bossard, C. C., J. M. Randall, and M. C. Hoshovsky, eds. 2000. Invasive plants of California's wildlands. Berkeley, CA: University of California Press, Berkeley. 360 p.

Broz, A. K., C. D. Broeckling, J. He, X. Dai, P. X. Zhao, and J. M. Vivanco. 2007. A first step in understanding an invasive weed through its genes: an EST analysis of invasive *Centaurea maculosa*. BMC Plant Biol. 7:25 http://www.biomedcentral.com/1471-2229/7/25.

Chao, W. S. 2008. Real-time PCR as a tool to study weed biology. Weed Sci. 56:290–296.

Chao, W. S., D. P. Horvath, J. V. Anderson, and M. P. Foley. 2005. Potential model weeds to study genomics, ecology, and physiology in the 21st century. Weed Sci. 53:929–937.

Downloaded From: https://bioone.org/journals/Weed-Science on 11 Oct 2024
Terms of Use: https://bioone.org/terms-of-use

Charmet, G., F. Balfourier, and V. Chatard. 1996. Taxonomic relationships and interspecific hybridization in the genus Lolium (grasses). Genet. Resour. Crop Evol. 43:319–327.

Church, S. A., K. Livingstone, Z. Lai, A. Kozik, S. J. Knapp, R. W. Michelmore, and L. H. Rieseberg. 2007. Using variable rate models to identify genes under selection in sequence pairs: their validity and limitations for EST sequences. J. Mol. Evol. 64:171–180.

Ciannamea, S., J. Busscher-Lange, S. de Folter, G. C. Angenent, and R.G.H. Immink. 2006. Characterization of the vernalization response in *Lolium perenne* by a cDNA microarray approach. Plant Cell Physiol. 47:481–492.

Devine, M. D. and A. Shukla. 2000. Altered target sites as a mechanism of herbicide resistance. Crop Prot. 19:881–889.

Dlugosch, K. M. and J. Whitton. 2008. Can we stop transgenes from taking a walk on the wild side? Mol. Ecol. 17:1167–1169.

Eckes, P., C. van Almsick, and M. Weidler. 2004. Gene expression profiling, a revolutionary tool in Bayer CropScience herbicide discovery. Pflanzenschutz-Nachr. Bayer 57:62–77.

Ellstrand, N. C. and K. A. Schierenbeck. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? Proc. Natl. Acad. Sci. U.S.A. 97:7043–7050.

Emberton, J., J. X. Ma, Y. N. Yuan, P. SanMiguel, and J. L. Bennetzen. 2005. Gene enrichment in maize with hypomethylated partial restriction (HMPR) libraries. Genome Res. 15:1441–1446.

Evans, G. M., H. Rees, C. L. Snell, and S. Sun. 1972. The relation between nuclear DNA amount and the duration of the mitotic cycle. Chromosomes Today 3:24–31.

Gardner, S. N., J. Gressel, and M. Mangel. 1998. A revolving dose strategy to delay the evolution of both quantitative vs. major monogene resistances to pesticides and drugs. Int. J. Pest Manag. 44:161–180.

Gressel, J. 2002. Molecular Biology of Weed Control. New York: Taylor and Francis. 504 p.

Gressel, J. 2008. Genetic Glass Ceilings. Baltimore: The Johns Hopkins University Press. 348 p.

Grime, J. P. 1977. Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. Am. Nat. 111:1169–1194.

Halfhill, M. D., L. L. Good, C. Basu, J. Burris, C. L. Main, T. C. Mueller, and C. N. Stewart, Jr. 2007. Transformation and segregation of GFP fluorescence and glyphosate resistance in horseweed (*Conyza canadensis*) hybrids. Plant Cell Rep. 26:303–311.

Heimann, B. and G. W. Cussans. 1996. The importance of seeds and sexual reproduction in the population biology of *Cirsium arvense*—a literature review. Weed Res. (Oxf.) 36:493–503.

Hieter, P. and M. Boguski. 1997. Functional genomics: it's all how you read it. Science 278:601–602.

Holm, L., J. Doll, E. Holm, J. V. Pancho, and J. P. Herberger. 1997. World Weeds: Natural Histories and Distribution. New York: John Wiley and Sons. 1152 p.

Horvath, D. P. and S. Clay. 2007. Heterologous hybridization of cotton (*Gossypium hirsutum*) microarrays with velvetleaf (*Abutilon theophrasti*) reveals physiological responses due to corn competition. Weed Sci. 55:546–557.

Horvath, D. P., J. V. Anderson, M. Soto-Suárez, and W. S. Chao. 2006a. Transcriptome analysis of leafy spurge (*Euphorbia esula*) crown buds during shifts in well-defined phases of dormancy. Weed Sci. 54:821–827.

Horvath, D. P., W. S. Chao, J. C. Suttle, J. Thimmipuram, and J. V. Anderson. 2008. Transcriptome analysis identifies novel responses and potential regulatory genes involved in seasonal dormancy transitions of leafy spurge (*Euphorbia esula* L.). BMC Genomics 9:536.

Horvath, D. P., R. Gulden, and S. A. Clay. 2006b. Microarray analysis of late-season velvetleaf (*Abutilon theophrasti*) effect on corn. Weed Sci. 54:983–994.

Horvath, D. P., R. Schaffer, and E. Wisman. 2003. Identification of genes induced in emerging tillers of wild oat (*Avena fatua*) using *Arabidopsis* microarrays. Weed Sci. 51:503–508.

Hudson, R. R., M. Kreitman, and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159.

Inderjit, R. M., R. Callaway, and J. M. Vivanco. 2006. Plant biochemistry helps to understand invasion ecology. Trends Plant Sci. 11:574–580.

Jofre-Garfias, A. E., N. Villegas-Sepúlveda, J. L. Cabrera-Ponce, R. M. Adame-Alvarez, L. Herrera-Estrella, and J. Simpson. 1997. *Agrobacterium*-mediated transformation of *Amaranthus hypochondriacus*: light- and tissue-specific expression of a pea chlorophyll a/b-binding protein promoter. Plant Cell Rep. 16:847–852.

Johnsen, Ø., C. G. Fossdal, N. Nagy, J. Mølmann, O. G. Daehlen, and T. Skrøppa. 2005. Climatic adaptation in *Picea abies* progenies is affected by the temperature during zygotic embryogenesis and seed maturation. Plant Cell Environ. 28:1090–1102.

Kane, N. C. and L. H. Rieseberg. 2007. Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. Genetics 175:1803–1812.

Kane, N. C. and L. H. Rieseberg. 2008. Genetics and evolution of weedy *Helianthus annuus* populations: evidence of multiple origins of an agricultural weed. Mol. Ecol. 7:384–394.

Karrenberg, S. and A. Widmer. 2008. Ecologically relevant genetic variation from a non-*Arabidopsis* perspective. Curr. Opin. Plant Biol. 11:156–162.

Kim, M., M-L. Cui, P. Cubas, A. Gillies, K. Lee, M. A. Chapman, R. J. Abbott, and E. Coen. 2008. Regulatory genes control a key morphological and ecological trait transferred between species. Science 322:1116–1119.

Lai, Z., B. L. Gross, Y. Zou, J. Andrews, and L. H. Rieseberg. 2006. Microarray analysis reveals differential gene expression in hybrid sunflower species. Mol. Ecol. 15:1213–1227.

Lai, Z., N. C. Kane, Y. Zou, and L. H. Rieseberg. 2008. Natural variation in gene expression between wild and weedy populations of *Helianthus annuus*. Genetics 179:1881–1890.

Landry, C. R., B. Lemos, S. A. Rifkin, W. J. Dickinson, and D. K. Hartl. 2007. Genetic properties influencing the evolvability of gene expression. Science 317:118–121.

Larrinua, I. M. and S. B. Belmar. 2008. Bioinformatics and its relevance to weed science. Weed Sci. 56:297–305.

Lee, R. M. and P. T. Tranel. 2008. Utilization of DNA microarrays in weed science research. Weed Sci. 56:283–289.

Lein, W., F. Börnke, A. Reindl, T. Ehrhardt, M. Stitt, and U. Sonnewald. 2004. Target-based discovery of novel herbicides. Curr. Opin. Plant Biol. 7:219–225.

Liu, X. S. 2007. Getting started in tiling microarray analysis. PLoS Comput. Biol. 3:e183.

Lokko, Y., J. V. Anderson, and S. Rudd, et al. 2007. Characterization of an 18,166 EST dataset for cassava (*Manihot esculenta* Crantz) enriched for drought-responsive genes. Plant Cell Rep. 26:1605–1618.

Mallory-Smith, C. and M. Zapiola. 2008. Gene flow from glyphosate-resistant crops. Pest. Manag. Sci. 64:428–440.

Malone-Schonenberg, J., C. J. Scelonge, M. Burrus, and D. L. Bidney. 1994. Stable transformation of sunflower using *Agrobacterium* and split embryonic axis explants. Plant Sci. 103:199–207.

Mardis, E. R. 2008. The impact of next-generation sequencing technology on genetics. Trends Genet. 24:133–141.

McDonald, J. H. and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351:652–654.

Michelmore, R., E. Marsh, S. Seely, and B. Landry. 1987. Transformation of lettuce Lactuca *sativa* mediated by *Agrobacterium tumefaciens*. Plant Cell Rep. 6:439–42.

Narumi, T., R. Aida, A. Ohmiya, and S. Satoh. 2005. Transformation of chrysanthemum with mutated ethylene receptor genes: mDG-ERS1 transgenes conferring reduced ethylene sensitivity and characterization of the transformants. Postharvest Biol. Technol. 37:101–110.

Neale, D. B. and P. K. Ingvarsson. 2008. Population, quantitative and comparative genomics of adaptation in forest trees. Curr. Opin. Plant Biol. 11:149–155.

Patzoldt, W. L., P. J. Tranel, and A. G. Hager. 2005. A waterhemp (*Amaranthus tuberculatus*) biotype with multiple resistance across three herbicide sites of action. Weed Sci. 53:30–36.

Pimentel, D., R. Zuniga, and D. Morrison. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecol. Econ. 52:273–288.

Prentis, P. J., J.R.U. Wilson, E. E. Dormontt, D. M. Richardson, and A. J. Lowe. 2008. Adaptive evolution in invasive species. Trends Plant Sci. 13:288–294.

Raghavan, C., E. K. Ong, M. Dalling, and T. W. Stevenson. 2005. Effect of herbicidal application of 2,4-dichlorophenoxyacetic acid in *Arabidopsis*. Funct. Integr. Genomics 5:4–17.

Raghavan, C., E. K. Ong, M. Dalling, and T. W. Stevenson. 2006. Regulation of genes associated with auxin, ethylene and ABA pathways by 2,4-dichlorophenoxyacetic acid in *Arabidopsis*. Funct. Integr. Genomics 6:60–70.

Rayburn, A. L., R. McLoskey, T. C. Jeschke, and P. J. Tranel. 2005. Genome size analysis of weedy *Amaranthus* species. Crop Sci. 45:2557–2562.

Renn, S. C., N. Aubin-Horth, and H. A. Hofmann. 2004. Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. BMC Genomics 5:42.

Sawbridge, T., E. Ong, and C. Binnion, et al. 2003. Generation and analysis of expressed sequence tags in perennial ryegrass (*Lolium perenne* L.). Plant Sci. 165:1089–1100.

Shendure, J. 2008. The beginning of the end for microarrays. Nat. Methods 5:585–587.

Stevens, P. F. 2007. Angiosperm Phylogeny. Version 8. http://www.mobot.org/MOBOT/research/APweb/. Accessed: January 13, 2008.

Downloaded From: https://bioone.org/journals/Weed-Science on 11 Oct 2024
Terms of Use: https://bioone.org/terms-of-use

Stewart, C. N., Jr., ed. 2009. Weedy and Invasive Plant Genomics. Ames, IA: Blackwell Scientific, 254 p.
Stinchcombe, J. R. and H. E. Hoekstra. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100:158–179.
Taji, T., M. Seki, and M. Satou, et al. 2004. Comparative genomics in salt tolerance between *Arabidopsis* and *Arabidopsis*-related halophyte salt cress using *Arabidopsis* microarray. Plant Physiol. 135:1697–1709.
Tajima, F. 1989. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. Genetics 123:85–595.
Tatematsu, K., S. Ward, O. Leyser, Y. Kamiya, and E. Nambara. 2005. Identification of *cis*-elements that regulate gene expression during initiation of axillary bud outgrowth in Arabidopsis. Plant Physiol. 138:757–766.
Thomas, M. A. and R. Klaper. 2004. Genomics for the ecological toolbox. Trends Ecol. Evol. 19:439–445.
Tranel, P. J. and F. Trucco. The *Amaranthus* complex: a model for weed genomics. *In* C. N. Stewart, Jr., ed. Weedy and Invasive Plant Genomics. Ames, IA: Blackwell, In press.
Tranel, P. J. and T. R. Wright. 2002. Resistance of weeds to ALS-inhibiting herbicides: what have we learned? Weed Sci. 50:700–712.
VanGessel, M. J. 2001. Glyphosate-resistant horseweed from Delaware. Weed Sci. 49:703–705.
Warwick, S. I. and C. N. Stewart, Jr. 2005. Crops come from wild plants: How domestication, transgenes, and linkage effects shape ferality. Pages 9–30 *in* J. Gressel, ed. Crop Ferality and Volunteerism. Boca Raton, Florida: CRC.
Weiss-Schneeweiss, H., J. Greilhuber, and G. M. Schneeweiss. 2005. Genome size evolution in holoparasitic *Orobanche* (Orobanchaceae) and related genera. Am. J. Bot. 93:148–156.
Whitelaw, C. A., W. B. Barbazuk, and G. Pertea, et al. 2003. Enrichment of gene-coding sequences in maize by genome filtration. Science 302:2118–2120.
Whitney, K. D., R. A. Randell, and L. H. Rieseberg. 2006. Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. Am. Nat. 167:794–707.
Wiehe, T., V. Nolte, D. Zivkovic, and C. Schlotterer. 2007. Identification of selective sweeps using a dynamically adjusted number of linked microsatellites. Genetics 175:207–218.
Wood, H. M., J. W. Grahame, S. Humphray, J. Rogers, and R. K. Butlin. 2008. Sequence differentiation in regions identified by a genome scan for local adaptation. Mol. Ecol. 17:3123–3135.
Wright, S. I. and B. S. Gaut. 2005. Molecular population genetics and the search for adaptive evolution in plants. Mol. Biol. Evol. 22:506–519.
[WSSA] Weed Science Society of America. 2008. Charting the course for weed genomics: the evolution of weediness. WSSA Annual Meeting, February 4–7, 2008, Chicago, Illinois.
Wu, C. A., D. B. Lowry, A. M. Cooley, K. M. Wright, Y. W. Lee, and J. H. Willis. 2008. Mimulus is an emerging model system for the integration of ecological and genomic studies. Heredity 100:220–230.
Wu, Y-Y., Q-J. Chen, M. Chen, J. Chen, and X-C. Wang. 2005. Salt-tolerant transgenic perennial ryegrass (*Lolium perenne* L.) obtained by *Agrobacterium tumefaciens*-mediated transformation of the vacuolar Na+/H+ antiporter gene. Plant Sci. 169:65–73.
Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. 15:568–573.
Yu, J., S. Hu, and J. Wang, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92.
Yuan, J. S., S. Y. Galbraith, S. Y. Dai, P. Griffin, and C. N. Stewart, Jr. 2008. Plant systems biology comes of age. Trends Plant Sci. 13:165–171.
Yuan, J. S., A. Reed, F. Chen, and C. N. Stewart, Jr. 2006. Statistical analysis of real-time PCR data. BMC Bioinformatics 7:85.
Yuan, J. S., P. J. Tranel, and C. N. Stewart, Jr. 2007. Non-target herbicide resistance: a family business. Trends Plant Sci. 12:6–13.
Zayed, A. and C. W. Whitfield. 2008. A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. Proc. Natl. Acad. Sci. U.S.A. 105:3421–3426.

Received January 14, 2009, and approved April 22, 2009.

Box 1. Tools in the Weed Genomics Toolbox

Bioinformatics. Bioinformatics refers to the handling and analysis of biological information using computers. It is the central tool in the genomics toolbox (Figure 2) because it is the means by which the large data sets inherent to genomic research are translated into biologically meaningful information. Numerous software packages are freely available or can be purchased for these purposes; however, some custom programming is almost always required to handle the data. Weed scientists entering genomics must either become comfortable with bioinformatics themselves or find ways to work closely with bioinformaticians or computer scientists (Larrinua and Belmar 2008). Good communication is imperative in these collaborations because the questions of interest to the researcher will guide how the data are acquired, organized, and analyzed.

Bioinformatics can include searching and acquisition of publicly available data, and extensive databases of gene sequences and expression patterns (such as Genbank and The Gene Expression Omnibus, both hosted by National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/) already exist, which could be a starting point for studying weed-related issues. Data obtained from genomic databases have been used by some weed scientists to design DNA microarrays for identifying global patterns of gene expression (Lai et al. 2008), for developing molecular markers for exploring evolution of biotypes (Kane and Rieseberg 2008), for mapping to identify genes responsible for specific phenotypes, and as resources for annotating sequences based upon similarity to model species and the burgeoning data from nonmodel species being deposited in databases.

Molecular Markers. Molecular markers are making a wide variety of contributions to weed science (Bodo Slotta 2008), and some types of these molecular markers are particularly suited to the high-throughput tracking of variation across a genome.

- *Amplified Fragment Length Polymorphisms (AFLPs)*. AFLPs determine variation in the length of randomly amplified regions of the genome after restriction fragmentation. Variation at a large number of loci is scored simultaneously by electrophoresis of a subset of fragments.
- *Single Nucleotide Polymorphisms (SNPs)*. SNPs determine variation in the nucleotide present at a single base-pair location. After identifying these loci through sequencing of multiple individuals, a wide variety of techniques are available for rapidly screening large numbers of loci and individuals.
- *Simple Sequence Repeats (SSRs)*: SSRs (also known as microsatellites) are regions of the genome where motifs of a few bases (two or three) are repeated a variable number of times. Polymorphism is apparent as variation of the length of the region, easily scored by electrophoresis after polymerase chain reaction (PCR) amplification of the SSRs.
- *Single Feature Polymorphisms (SFPs)*: Sequence variation is revealed by variation among individuals in the ability of their DNA to hybridize to a particular location on a tiling array (see below).

Molecular Maps. Molecular maps describe how molecular markers are positioned relative to one another in the genome and how individual markers or sets of markers relate to phenotypes. These allow us to understand the genetic basis of weediness traits and how genome structure and its evolution have affected weed biology.

Downloaded From: https://bioone.org/journals/Weed-Science on 11 Oct 2024
Terms of Use: https://bioone.org/terms-of-use

- *Genetic Maps*. Genetic maps show the relative positions (order) of genetic markers in the genome, as determined by their linkage (recombination distances) to one another.
- *Quantitative Trait Loci (QTLs)*. QTLs are genomic regions associated with a particular phenotypic effect, as defined by a set of marker loci, which correlate with the phenotype. QTL mapping relies upon controlled crosses to generate known relationships among individuals, which yield known probabilities that markers are identical by descent and which permit their association with shared phenotypes among individuals.
- *Association Mapping*. Association mapping allows identification of QTLs in individuals with unknown pedigree by using genetic similarities to infer probability of identity by descent. Because genotypes used for association mapping are less closely related than those used for QTL mapping, much finer-scale mapping is typically possible.
- *Physical Maps*: The physical distances (bp) between genetic markers are shown in physical maps. These may include full sequence information for a region, for instance through the use of bacterial artificial chromosomes (BACs), where regions of several thousands of base pairs are cloned into bacteria and fully sequenced.
- *Map-Based Cloning*. Map-based cloning allows identification of a candidate gene of interest through physical mapping and sequencing within a region defined by markers in a genetic map.

DNA Sequences. DNA sequences generate direct insights into the genetic makeup and evolution of weeds, as well as indirect information useful for most of the genomic tools listed here.

- *Whole-Genome Sequencing*. The sequencing of the genome in its entirety, including noncoding regions as well as transcribed and untranscribed coding regions provides a level of detail that facilitates assembly of resequencing data for population genomics, discovery and annotation of coding regions, identification of promoter sequences for particular genes of interest, and development of molecular maps, map-based cloning, and marker development. These data will be particularly integral to the identification of common regulatory elements from promoters associated with clusters of coordinately expressed genes.
- *Expressed Sequence Tags (ESTs)*. ESTs are sequences of transcribed genes (mRNAs) present in a particular sample or pool of samples. ESTs are an efficient way to focus in on the transcribed portion of the genome and can be used to study gene evolution, to obtain molecular markers and variation for population genomics, and to generate probes for microarray development.
- *Gene Space Sequencing*. Gene space sequencing is sequencing of the low-copy, gene-rich portion of the genome (Barbazuk et al. 2005). A filtration approach is used to enrich a sample for gene space and then the sample is sequenced (Emberton et al. 2005; Whitelaw et al. 2003). Gene space sequencing is similar to expressed sequence tagging but with a higher proportion of genes discovered, particularly large genes and those that are weakly or rarely expressed (e.g., transcription factors and disease resistance genes). Gene space sequencing also yields information on promoters, introns, and other nonexon sequences that are critical to analyses of gene structure and evolution.
- *Next-Generation Sequencing (NGS)*. NGS methods are recently developed, high-throughput alternatives to traditional sequencing, offering simultaneous sequencing of hundreds of thousands of short regions of a DNA sample (Mardis 2008). Traditional (Sanger) sequencing returns a single sequence of up to about 1,000 bp/sample. NGS methods involve finer fragmentation of a large sample of DNA, distribution of those fragments across a slide or the plate of microscopic wells, and simultaneous sequencing of the fragments. These techniques can be used for any of the above genomic-sequencing approaches, at a significantly reduced cost per project relative to Sanger sequencing, although shorter read lengths limit their ability to provide complete de novo sequencing of repetitive (low complexity) regions of the genome. NGS methods can also be performed such that the number of reads for a particular sequence correlates with its frequency in the sample; sequencing of a library of expressed sequences in this way provides both sequence and a measure of expression levels.

Microarrays. Microarrays are slides printed with a set of cDNA or oligonucleotide probes, to which corresponding sequences in a sample of expressed sequences (a *transcriptome*) will anneal (*hybridize*). Samples are fluorescently labeled, such that hybridization intensity correlates with expression level for a single sample or color distinguishes relative expression levels when two samples are labeled with different dyes and pooled to compete for the same sites on an array. Microarrays allow efficient comparisons of transcriptomes across species, populations, tissues, or plants grown under various conditions (see Lee and Tranel 2008).

It has been suggested that NGS (above) might render microarrays obsolete (Shendure 2008). However, because of library construction costs for sequencing, microarrays are still the most cost-effective method. High-throughput companies (NimbleGen, Agilent, Affymetrix, etc.) now offer to generate the arrays at little or no cost and provide a fixed-fee schedule to run the experiments and supply the scientist with transcriptome expression data.

- *cDNA Arrays*. cDNA array probes are PCR-amplified clones from a cDNA library, with probes up to a few thousand base pairs in length. These long probes offer the potential for hybridization across different species with divergent homologous sequences, making the arrays useful for multiple species and for direct cross-species, competitive hybridization experiments.
- *Oligo Arrays*. Oligo array probes are synthesized oligonucleotides. Long probes (about 70 bp) are similar to cDNA arrays but offer increased accuracy and reproducibility. Short-probe arrays (about 20 bp) are available for single-sample experiments and offer higher specificity, with the potential to differentiate among members of gene families.
- *Tiling Arrays*. Tiling array probes are short oligonucleotides designed to cover the entire genome or contigs of interest. Depending on the experiment and the length and overlap of the probes, tiling arrays can be used to examine details of expression variation, transcription factor binding sites, copy number variation, or DNA methylation, or to map transcriptomes in sequenced genomes (Liu 2007).

Downloaded From: https://bioone.org/journals/Weed-Science on 11 Oct 2024
Terms of Use: https://bioone.org/terms-of-use

Functional Characterization. Functional characterization involves a set of tools designed to help us understand the function, regulation, and phenotypic effects of particular loci and alleles.

- *Real-Time Reverse-Transcriptase PCR (real-time RT-PCR)*. Real-time RT-PCR quantifies PCR products by florescence after each amplification cycle, also known as *quantitative PCR*. This permits direct assessment of the quantity of a particular mRNA transcript in the tissue of interest (see Chao 2008; Yuan et al. 2006).
- *Transgenic Overexpression*. Transgenic insertion of a gene and its promoter into an individual induces increased expression of that gene, demonstrating the phenotypic effects of its up-regulation.
- *Gene Knockdowns*. Gene knockdowns are reductions in gene expression through either genetic modification or treatment with an oligonucleotide that interferes with gene or mRNA function. These demonstrate the effects of down-regulation or loss of function of the gene of interest.
- *Mutagenesis*. Induction of mutations demonstrates the effects of allelic variation or loss of function in a gene.

Downloaded From: https://bioone.org/journals/Weed-Science on 11 Oct 2024
Terms of Use: https://bioone.org/terms-of-use