Review TRENDS in Plant Science Vol.9 No.8 August 2004

Weed genomics: new tools to understand weed biology

Chhandak Basu, Matthew D. Halfhill, Thomas C. Mueller and C. Neal Stewart Jr

Department of Plant Sciences, 2431 Joe Johnson Drive, University of Tennessee, Knoxville, TN 37996, USA

In spite of the large yield losses that weeds inflict on crops, we know little about the genomics of economically important weed species. Comparative genomics between plant model species and weeds, map-based approaches, genomic sequencing and functional genomics can play vital roles in understanding and dissecting weedy traits of agronomically important weed species that damage crops. Weed genomics research should increase our understanding of the evolution of herbicide resistance and of the basic genetics underlying traits that make weeds a successful group of plants. Here, we propose specific weed candidates as genomic models, including economically important plants that have evolved herbicide resistance on several occasions and weeds with good comparative genomic qualities that can be anchored to the genomics of *Arabidopsis* and *Oryza sativa*.

It is no secret that weeds are farmers' worst pest, causing up to US$20 billion in losses every year in the USA alone [1], even though there are a plethora of cultural practices and chemical applications to limit weed damage. Worldwide, between 67 and 104 plant taxa are responsible for 90% of the economic damage caused by weeds, so most crop losses are caused by few species [2,3]. Although *Arabidopsis* is commonly referred to as a weed, it does not cause any significant economic loss and thus it is a poor weed [4]. Weed scientists have diverse definitions for weeds, the simplest of which is 'a plant growing out of place'; therefore, a corn plant is considered to be a 'volunteer' weed when growing in a soybean field. A more precise definition might reflect the economic losses stemming from weed problems in agriculture. The kinds of weeds that agriculturalists are most concerned with are highly competitive plants that decrease the growth and yield of crops. Thus, we refer to weeds here as plants that are unusually persistent and pernicious, that significantly interfere with the growth of crops, and that are optimally adapted to agroecosystems [5]. More often than not, major weeds are inadvertently domesticated species that do not flourish outside of agriculture. Thus, without the farmer, weeds would not exist as we know them.

Why weed genomics?

Genomics is the study of the structure, function and evolution of genomes. Given the economic importance of weeds, it is surprising that we know relatively little about the genomics of weediness [6]. Transgene flow from crops to weeds is also a concern, through which weed species might gain novel genetic characteristics from transgenic crops (Box 1). A deeper understanding of the genomics of weed species is needed to design better weed control. FUNCTIONAL GENOMICS (see Glossary) and COMPARATIVE GENOMICS are valuable research approaches by which we can investigate the competitive abilities of weedy plants [1,4,7,8]. However, genomics has been little exploited for this purpose. It is urgent that weed genomic models be developed – the next logical step up the stairway of plant genomics.

In this article, we analyse our state of knowledge of weed genomics and propose incremental steps to increase genomics research in WEED BIOLOGY. The overarching goal for any functional weed genomics program is the discovery and characterization of weediness genes. The characterization of weediness alleles would be one focus of such research because few gene families are expected to be truly novel to weed species [9]. But this might prove to be naive because weeds could have novel transcription factors, new regulatory pathways and unique homeotic genes that allow them to respond and grow differently than crops and wild plants not adapted to agroecosystems. We do not know which of these factors would be most important and so functional and comparative genomic approaches are needed that take advantage of existing plant genomic knowledge and techniques. The current view of plant genomics is one sided, being tilted towards

Glossary

Comparative genomics: relative analyses of entire sets of genes from different organisms.
Fecundity: reproductive capacity.
Functional genomics: understanding the roles of all genes, promoters and regulatory elements of an organism.
Forward genetics: the classical genetics approach in which a mutant phenotype is first identified and the gene responsible for the mutant phenotype is then cloned and characterized.
Reverse genetics: starting with a sequence of a mutant gene and then observing the phenotype resulting from the mutant gene.
Expressed sequence tags: parts of transcripts that are generated from a subset of cDNAs.
Microarray: cDNAs or oligonucleotides (representing whole or partial genome of an organism) immobilized on a glass slide or other substrate and probed with labelled cDNAs from treated and control tissues for gene expression analysis.
Monoecious: individual flowers or plants containing both male and female reproductive parts.
Weed biology: the science of studying weed morphology, growth, evolution and interaction with crop species.

Corresponding author: C. Neal Stewart (nealstewart@utk.edu).
Available online 17 July 2004

 1360-1385/$ - see front matter doi:10.1016/j.tplants.2004.06.003

Box 1. Gene flow and introgression from transgenic crops to weeds

A thorough review of pertinent literature relevant to the topic of transgene introgression from crops to weeds has recently been published [32]. The movement of herbicide tolerance traits from crops to related weeds has been a fear that has spawned significant regulation and scientific study. The spontaneous evolution and selection of herbicide tolerance *de novo* in weeds has caused new weed problems [4]. Gene flow from transgenic crops to weeds is possible, and we present a brief case-by-case analysis of potentially problematic crops and weeds. We recognize [32] that herbicide tolerance is a trait for crop-to-weed introgression of special importance in agriculture and list the following crop–weed combinations in the moderate risk category:

- Alfalfa (*Medicago sativa* ssp. *sativa*) with wild alfalfa (*Medicago sativa*).
- Bread wheat (*Triticum aestivum*) and durum wheat (*Triticum turgidum* ssp. *turgidum*) with wild *T. turgidum* and *Aegilops* species.
- Canola (*Brassica napus*) with wild turnip (field mustard, *Brassica rapa*; to confuse the matter slightly, there is also canola-quality *B. rapa*).
- Sunflower (*Helianthus annuus*) with wild sunflowers (*Helianthus* species).

The single combination in the high-risk category is sorghum (*Sorghum bicolor*) with johnsongrass (*Sorghum halepense*). This last combination is of special interest because of the local noxiousness of johnsongrass. It is unlikely that any transgenic crop sorghum will be engineered, and certainly not for herbicide tolerance.

The other combination of special interest worth noting here is canola with wild turnip. Hybridization in an agronomic situation has occurred between these two plants in the field in Canada, where herbicide-tolerant canola varieties have been extensively planted [64]. Thus, the first crop×wild hybrids for glyphosate tolerance have arisen spontaneously, with the potential for backcross hybrids (back to wild *B. rapa*) now a reality. However, these new transgenic herbicide-tolerant hybrids are on small plots of land, whereas there are hundreds of thousands of hectares containing glyphosate-tolerant horseweed (*Conyza canandensis*) in the USA that have arisen via spontaneous mutation and selection, and not from gene flow from transgenic crops [65]. Thus, spontaneous evolution of herbicide resistance would appear, at first glance, to be orders of magnitude more important than introgressed herbicide tolerance from crops to related weeds, yet this is barely recognized by regulators and high-profile research.

crops. In a small way, we hope that this review will begin to remedy that situation. Although we might not marry weed science with genomics here, perhaps the two will at least begin a conversation.

Weediness characteristics

Weed science has given us a thorough knowledge of weedy traits (Table 1) but we are largely ignorant about the functional genomics underlying them. Using the available tools of genomics, we can improve our understanding of weed biology by finding and characterizing genes that might play a role in fitness, competitiveness and adaptations of weeds in agroecosystems [10]. Hence, weed genomics could integrate the concept of functional specificity in an evolutionary context by pinpointing which genes are responsible for the specific traits that characterize the weed niche in agriculture. Just as there are wild fungi and pathogens, so there are wild plants and weeds – and, in both cases, the pathogens and weeds have virulence factors, which are the target of genomics discovery. The selection of weed species for study, the development of screening strategies and the functional

Table 1. Description of weediness characteristics prevalent in most weeds[a]

Weediness characteristic	Description of the characteristic
Rapid seedling growth	Allows maximum capture of growth-limiting factors, such as light, water and nutrients.
Short vegetative phase	Under situations that cause a plant to germinate later in the growing season, a plant can still complete its life cycle and produce viable seed.
Indeterminant	Flowering throughout the season, while growing vegetatively over an extended period of time.
Self-compatible	Allows genetic divergence from previous generations, but the plants do not require special pollinators such as bees or other insects to produce viable seed.
High seed output	Plants produce many seeds that are dispersed both spatially and temporally into as many favourable locations as possible for subsequent growth. Horseweed (*Conyza canadensis*) plants produce more than 100 000 seeds per plant.
Long-distance seed dispersal	Given that terrestrial plants are not mobile, the only chance for long-distance spread is via seed or other propagules. Seeds that float in water or are carried by the wind are more widely distributed. Dandelion seeds can be moved long distances by wind.
Competitive with crop plants	Weeds can compete for light, nutrients and water, thereby substantially reducing crop yield. Average of 54 foxtail (*Setaria faberii*) plants per foot (~0.3 m) of row can reduce grain yields of soybean and corn to 24% and 28% of the original levels, respectively [69].
Deep root system	Deep root systems allow weeds to thrive in drought conditions. Field bindweed (*Convolvulus arvensis*) roots can penetrate up to 10 feet (~3 m) deep into the soil.
Discontinuous dormancy	Weed seeds might be dormant in unfavourable environmental, ensuring germination when environmental conditions are again favourable. Common lambsquarters (*Chenopodium album*) is a weed that displays extensive dormancy mechanisms.
Environmentally plastic	A plant can change its growth form in response to environmental factors or other control strategies. Goosegrass (*Eleusine indica*) has an upright growth form under field conditions, but it develops a prostrate shape when it is mowed (such as on a golf course). Both types of plants can produce viable seed.
Dual modes of reproduction	Weeds can reproduce sexually and asexually. Although most of them reproduce sexually and from seeds, many weeds can reproduce asexually [e.g. common bermudagrass (*Cynodon dactylon*) can spread asexually by forming stolons, and also sexually by forming seeds].
Allelopathy	Plants produce chemicals that discourage the growth or kill other plants.

[a]Adapted from Refs [50,67,68].

analysis of weediness traits are required in the conception of weed genomics models.

Weed genomics to date

Quantitative traits

Molecular markers of various types (e.g. random amplified polymorphic DNA, amplified fragment length polymorphisms and simple sequence repeats [11,12]) have been used to study genetic diversity and population biology, as well as to identify quantitative trait loci (QTLs) in weed species. QTL mapping aims to identify regions in the genome that affect quantitative traits and is one logical entry point into weed genomics. Many weedy traits (such as rapid plant growth [13], dormancy [14,15] and competitiveness [16]) are relevant quantitative traits (Table 1). The following are a few illustrative examples. An intraspecific hybrid population was generated through crossing of weeds *Senecio vulgaris* var. *vulgaris* with *S. vulgaris* ssp. *denticulatus* (the ancestor of *S. vulgaris* var. *vulgaris*) [17]. After hybrids were selfed, the constructed linkage map was used to identify three chromosomal regions responsible for controlling seed dormancy, speed of development and the presence or absence of ray florets (traits that distinguished the two subspecies). In another study of seed dormancy, wild oat (*Avena fatua*) recombinant inbred lines were screened for seed dormancy and two loci associated with early germination were identified, as was one locus associated with late germination [18]. Johnsongrass (*Sorghum halepense*) is a particularly noxious weed in the USA. Three QTLs associated with vegetative dispersal by rhizomes, a weediness trait, were predicted in johnsongrass [19]. Progeny resulting from a cross between cultivated and wild species of *Sorghum* (including a progenitor of johnsongrass, a hybrid) were screened using maps and important QTLs were identified. In another study, major QTLs that were responsible for perennial habit were identified in johnsongrass and rice (*Oryza sativa*), indicating convergence in these traits [20]. Finally, a QTL for competitiveness was identified in wheat (*Triticum aestivum*) when grown against the weed annual ryegrass (*Lolium rigidum*) [16].

Microarrays

MICROARRAY analysis, a functional genomics tool used for studying transcriptional regulation, has recently been explored in weed science. An *Arabidopsis* microarray was used to investigate differential gene expression in underground buds of leafy spurge (*Euphorbia esula*) to examine dormancy [21]. The researchers identified more than 16 differentially expressed genes in underground buds of leafy spurge during dormancy break. Tillering of wild oat (*A. fatua*) was studied using *Arabidopsis* microarrays [22]. David Horvath *et al.* used *Arabidopsis* microarrays to investigate transcriptional regulation in both wild oat and leafy spurge [23]. Although researchers are investigating the degree of power in microarray analysis, and microarray resources and affordability are ever increasing, it is worthwhile noting that heterologous probing from a wide range of taxa (e.g. from monocot cDNAs to dicot arrays) will probably be problematic because of the divergence of all but the most-conserved genes. Another limitation of microarray analysis is that weediness genes might not be significantly upregulated (e.g. twofold) and so results would not be useful.

Selection of weed genomic model species

Genomics research performed on the ultimate model weed, *Arabidopsis*, has been successful [24]. Even though *Arabidopsis* is a poor weed, it is an excellent model plant because of its small size, small genome, quick generation time and ease of genetic transformation [25]. Rice [26,27] has emerged as the first model genomic crop and it, too, possesses many of these qualities. Thus, we would emulate these successes in arguing for the community to establish certain model genomic weeds (Table 2). In developing this argument, we shall also discuss some of the methods that might help to identify weediness genes.

The competitive ability of agronomic weeds is rooted in the interplay of weediness characteristics (Tables 1,3). Some weeds simply have substantial early season growth that confers an advantage in withstanding the farmer's best efforts to eradicate them. For most weeds, prolific seed production and dormancy allow indefinite survival of populations [14]. Selection of weed species candidates for future genomic studies is crucial, and is determined by the intersection of appropriate screenable weedy traits and tractable genomic characterizations (Table 3).

For our purposes here, we focus solely on agronomic weeds, not perennials. Thus, we are ignoring many important exotic invasive species that can disrupt rangeland or natural ecosystems. That is not to say that invasive species are not appropriate for genomic study, but they are outside of the scope of this review. Weeds of interest are primarily crop mimics that compete with crops in identical niches.

Amaranthus species (also known as pigweeds), horseweed (*Conyza canadensis*), wild turnip (*Brassica rapa*) and weedy red rice (*O. sativa*) are all potential candidates for weed genomics studies. *Amaranthus* spp. are vigorous summer annuals that have widespread agronomic importance in crops. Common waterhemp (*Amaranthus rudis*) is a dominant weed in the midwestern USA, infesting several million hectares [28]. This taxon is particularly interesting because it has adapted from wetland plant to important weed in a matter of decades. Palmer pigweed (*Amaranthus palmeri*) is a dominant weed covering millions of hectares of the southern USA. Smooth pigweed (*Amaranthus hybridus*) might be the best pigweed genomic model because, unlike the others, it is MONOECIOUS and self-compatible.

The second model weed candidate, *C. canadensis*, is a winter annual that has adapted by being able to germinate under various environmental conditions. It is a cosmopolitan weed that has quickly evolved resistance to a wide range of herbicides, the latest being glyphosate (e.g. Roundup®) [29,30]. This adaptive ability to overcome herbicide control warrants further examination to elucidate the underlying molecular factors of resistance (Box 2; http://www.weedscience.org/). *Amaranthus* spp. and *C. canadensis* should ultimately be studied using the full suite of genomic tools but are probably not the appropriate

Table 2. Candidate weed species that are potential models for weed genomics. *Arabidopsis*, rice and maize are included for comparisons

Weed species	DNA content (Mbp per haploid cell)[a]	Chromosome number (2*n*)	Ploidy level	Plant size[b]
Common waterhemp (*Amaranthus rudis*)	657 [70]	?	?	Large
Smooth pigweed (*Amaranthus hybridus*)	686	?	?	Large
Palmer pigweed (*Amaranthus palmeri*)	?	?	?	Large
Pitted morning glory (*Ipomoea lacunose*)	735	30	2	Medium
Horseweed (*Conyza canadensis*)	200–600[c]	18 [71]	2 [71]	Large
Wild turnip (*Brassica rapa*)	784	20	2	Medium
Common lambs quarter (*Chenopodium album*)	2279	54	6	Large
Prickly sida (*Sida spinosa*)	956	?	?	Medium
Dandelion (*Taraxacum officinale*)	1250	24	3	Small
Johnsongrass (*Sorghum halepense*)	1617	40	4	Large
Purple nutsedge (*Cyperus rotundus*)	441	?	?	Small
Giant foxtail (*Setaria faberi*)	1593	36	4	Medium
Large crabgrass (*Digitaria sanguinalis*)	1176	54	6	Medium
Annual ryegrass (*Lolium multiflorum*)	4018	14	2	Medium
Thale cress (*Arabidopsis thaliana*)	172	10	2	Small
Rice (crop and weed) (*Oryza sativa*)	490	24	2	Medium
Maize (crop) (*Zea mays*)	2761	20	2	Large

Abbreviation: ? indicates unknown.
[a]All DNA content, chromosome number and ploidy level data and references cited in this table obtained from http://www.kew.org/cval, except for tall waterhemp and horseweed.
[b]Plant sizes: large, plant >200 cm high; medium, plant between 30 cm and 200 cm high; small, plant less than 30 cm high.
[c]Flow cytometry was performed by the authors of this paper according to established protocol [72].

first choices for dissecting weediness using genomics because of the cost of genome-wide sequencing. However, if funds were available, we recommend that the first full-blown genomics project on a weed be performed using an *Amaranthus* sp. because of their rapid evolution, widespread dominance and economic importance. It is the consummate weed.

Wild turnip (*B. rapa*) and red rice are potentially the best models because there is an inordinate amount of latent genomics information within arm's reach. Wild turnip, a true diploid, shares the AA genome with canola (*Brassica napus*, AACC) and is an important recipient of herbicide resistance and other transgenes introgressed from canola [31,32]. Most importantly, it is a close relative of *Arabidopsis* in the Brassicaceae. As a result of extensive synteny, much of the *Arabidopsis* genomics would be directly relevant to *Brassica* [33,34] (for more information on the *Brassica* genome initiative, see http://www.brassica.info/). In addition, there are many agronomically important *B. rapa* crop types (e.g. canola, Polish rape, turnip) that could be compared with weedy *B. rapa* types.

Finally, weedy red rice is usually the same species as cultivated Asian rice [35], which is a genomics model in its own right. Although important *Oryza* species have several genomes, *O. sativa* types (red and cultivated) share the AA genome (different to the AA genome of *Brassica*). With regard to weediness and domestication, rice research is probably the most advanced and is discussed below. By first investigating wild turnip and red rice for weediness genes, sequence data, microarrays and other genomic

Table 3. Weediness characteristics and weed species of potential interest. *Arabidopsis*, rice and maize are included for comparisons

Weediness characteristic	Weed species														Crop and model species		
	Amaranthus rudis	*Amaranthus hybridus*	*Amaranthus palmeri*	*Ipomoea lacunosa*	*Conyza canadensis*	*Brassica rapa*	*Chenopodium album*	*Sida spinosa*	*Taraxacum officinale*	*Sorghum halepense*	*Cyperus rotundus*	*Setaria faberi*	*Digitaria sanguinalis*	*Lolium multiflorum*	*Arabidopsis thaliana*	*Oryza sativa*	*Zea mays*
Rapid seedling growth	+	+	+	+	−	−	+	+	+	+	+	+	+	+	+	−	−
Short vegetative phase	+	+	+	−	+	−	+	+	+	−	−	−	+	−	+	−	−
Indeterminant	+	+	+	+	+	+	+	?	+	+	+	+	+	+	−	−	−
Self-compatible	−	+	−	+	+	−	+	+	−	+	+	+	+	+	+	+	+
High seed output	+	+	+	−	+	+	+	−	+	+	−	+	−	+	+	−	−
Long-distance seed dispersal	+	+	+	−	+	−	−	+	+	+	−	−	−	−	−	−	−
Competitive with crop plants	+	+	+	+	−	+	+	+	−	+	+	+	+	+	−	−	+
Deep root system	−	−	−	−	+	−	−	−	+	+	+	−	−	−	−	−	−
Discontinuous dormancy	+	+	+	+	+	+	+	+	+	+	+	+	+	?	−	−	−
Environmentally plastic	−	−	−	−	+	?	+	+	+	+	+	+	+	+	−	−	−
Dual mode of reproduction	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−	−	−
Allelopathy	−	−	−	−	+	?	+	−	−	+	+	+	+	−	−	−	−

Abbreviations: +, present; −, absent; ?, unknown.

Box 2. Herbicide resistance in weeds

Herbicide resistance is a weed physiological characteristic that has the potential to cause significant economic losses. There are 286 resistant biotypes and 171 species (102 dicots and 69 monocots) of weeds have been reported to be herbicide resistant from >270 000 fields worldwide (http://www.weedresearch.com/). As defined by the Weed Science Society of America, herbicide resistance is 'the inherited ability of a plant to survive and reproduce following exposure to a dose of herbicide normally lethal to the wild type. In a plant, resistance can be naturally occurring or induced by such techniques as genetic engineering or selection of variants produced by tissue culture or mutagenesis (http://www.weedresearch.com/). The ten most economically important herbicide resistant weeds are (http://www.weedresearch.com/):

- Horseweed (*Conyza canadensis*)
- Redroot pigweed (*Amaranthus retroflexus*)
- Rigid ryegrass (*Lolium rigidum*)
- Wild oat (*Avena fatua*)
- Smooth pigweed (*Amaranthus hybridus*)
- Barnyardgrass (Echinochloa crus-galli)
- Goosegrass (*Elusine indica*)
- Common lambsquarters (*Chenopodium album*)
- Kochia (*Kochia scoparia*)
- Green foxtail (*Setaria viridis*)

The mechanism of herbicide resistance is a multifaceted molecular genetic phenomenon. The s-triazine herbicides inhibit the electron transport chain in photosystem II [50]. It has been reported that mutation in the triazine-binding site in the chloroplast produces triazine-resistant weeds (e.g. *Amaranthus* spp.) [50]. The sulfonylurea and imidazolinone herbicides inhibit the production of acetolactate synthase (ALS), an enzyme controlling the biosynthesis of branched-chain amino acids [4,50]. Sulfonylurea and imidazolinone-herbicide-resistant weeds produce a modified ALS that is incapable of intercalating with the herbicides and thereby negates herbicide action [50]. The glyphosate-containing herbicides inhibit 5-enolpyruvylshikimate 3-phosphate synthase (EPSPS), which is involved in the production of aromatic amino acids. Apparently, *C. canadensis* has three EPSPS genes [66]. One of these has the P106T active site change that one would predict would confer resistance. However, this change was found in both susceptible and resistance biotypes. It has also been determined that resistance exhibits a 3:1 segregation pattern in test crosses [66]. These data suggest that a mutation in a single gene is responsible for resistance – that gene is not one coding for EPSPS. Another *Conyza* sp. (*Conyza bonariensis*) has evolved glyphosate tolerance in South Africa. Although unconfirmed, some scientists also propose that the ATP-binding cassette (ABC) transporters, which play roles in ion exchange and sometimes detoxification of plant cells, might play roles in weeds that could make them herbicide resistant [4]. The evolution and mechanism of herbicide resistance in weeds is discussed in Ref. [4].

tools might be immediately applied to understand better the molecular basis of weediness [26,27].

Genomic approaches to dissect weediness traits

Segregation analysis and molecular mapping

Segregation analysis and molecular mapping of populations are useful tools to identify genes associated with weediness traits. In cultivated and weedy biotypes of *B. rapa* and *O. sativa*, the segregation of weediness or domestication traits could be screened in F_2 populations. Comparative mapping would be useful to analyse QTLs inherited from parents. A straightforward alignment of available maps could be used to find genomic regions that contribute to weediness. Such an approach has been fruitful in rice, in which domestication and weediness traits were mapped to a few chromosomal regions [36]. Loren Rieseberg *et al.* [37] advocated the use of map-based approaches to study the evolution of wild plants, an argument that works equally well for studying weeds. A weedy rice biotype with increased seed dormancy was crossed with a crop breeding line to produce BC_1 individuals (backcrossed generation one) and QTLs for dormancy were identified [38]. Six QTLs explaining significant dormancy and epistases led the researchers to infer that dormancy in rice has complex control networks, which might have important implications for genomics research in weeds.

Sequencing and comparative genomics

Genomic regions of interest identified by QTL and sequence information from other model plant species can be used to assess candidate weed genes. The rice genome has already been sequenced [26,27]. Likewise, *B. rapa* weed and crop types could also be sequenced. Although the cost would be significant, the genomes of several weed species representing diverse taxa might be also be sequenced [39]. After shotgun sequencing [40] of model weed species, nucleotide information could be compared with that from the *Arabidopsis* and rice projects as well as other known sequences in available databases. Sequence homology of weediness genes and those of known function would reveal information about genomic structure and function, gene interactions, species relatedness, post-translational events and more.

Comparative genomics approaches are not limited to Brassica and Oryza weeds

Jointed goatgrass (*Aegilops cylindrical*, CCDD) and wheat (*T. aestivum*, AABBDD) share one common genome of the four that compose the two species [41] and so there is interest in using wheat genomics (to date, mainly in the form of EXPRESSED SEQUENCE TAGS and maps) to identify post-harvest sprouting alleles in the weed [14,42]. In another study, rice genomics has been useful in studying dormancy-specific cDNA clones *Ecd1 and Ecd2* (the clones were originally identified from *Echinochloa* spp., weeds in rice and other crops) [43]. The clones were more upregulated in seeds of the dormant biotype than in the non-dormant biotype of *Echinochloa* spp. The clone *Ecd2* had significant homology with the α chain of mitochondrial H^+-transporting ATP synthase in rice. This information has been helpful in elucidating the mechanism of *Echinochloa* seed dormancy in the field (e.g. by aerobic respiration and ATP production). If such a heterologous intergenic approach has been successful then an intraspecific approach to dissect weediness in a weed such as red rice should be even more powerful. We believe a comparative genomics approach is a good candidate to yield a high data-to-cost ratio.

Forward genetics: from phenotype to gene

Screening mutant plant populations for the identification of genes is a classical FORWARD GENETICS approach and has

been successfully used for years in many plant species. Two T-DNA insertional mutagenesis strategies [activation tagging (upregulation of genes) [44] and gene knockout libraries (disrupting gene functions) [45]] could be used in an effective forward genetics screens for weediness traits if high-throughput transformation were to be successfully developed in weeds. The lack of high-throughput genetic transformation technologies currently limits the production of mutant libraries in rice. Although traditional tissue-culture-based transformation systems can be developed for weeds [46] [as we have recently done for horseweed and wild turnip (C. Basu *et al.*, unpublished)], it would be advantageous to develop *in planta Agrobacterium tumefaciens*- mediated transformation to engineer egg cells directly on the plant [47]. T-DNA insertional mutagenesis has become routine in *Arabidopsis* using the *in planta* floral dip method [48]. Because weeds typically produce many seeds, there seems to be the potential to develop *in planta* transformation in such plants. Indeed, *B. napus*, a close relative of wild turnip, has successfully been transformed by the floral dip method [49].

With the development of a high-throughput transformation system, weediness screens could be performed sequentially: seed dormancy, rapid early season growth, flowering time, potential for abiotic stress tolerance and FECUNDITY. As an example, dormancy is an important weediness trait: weed seeds might remain dormant for many years [50]. Genetic control of seed dormancy has already been described, so there is some background information available now. Dormancy loci in wild oat (*A. fatua*) have been previously identified [18]. A mutant weed population could be screened and dormancy mutants could be identified by selecting the plants lacking dormancy. Similarly, fecundity genes have been identified in other species [48]. Fecundity mutants might be selected by identifying plants producing few seeds. Flowering time has a genetic basis [51–54], and early- and late-flowering mutant plants could be further characterized for important genes controlling weed flower phenology. Herbicide-resistant weeds can easily be identified using a herbicide screen (survivors are resistant). Often, farmers discover herbicide-resistant biotypes surreptitiously in agricultural fields following herbicide application.

Even though transformation methods could be helpful in assaying weediness genes, epistasis and redundancy could be problematic. For instance, Xing-You Gu *et al.* [38] found that knocking out one of six genes controlling dormancy would predictably result in nearly no phenotype change because dormancy is a quantitative trait. Potentially, it might be necessary to breed near-isogenic lines for a knockout approach to be informative (compare with. Ref. [55]).

Reverse genetics: from genes to phenotype

DNA sequencing and advanced genomics have yielded innumerable genes with unknown functions. REVERSE GENETICS is an approach that relies on mutating specific genes and then observing the resulting phenotypes. Alternatively, the gene can also be overexpressed under the control of a strong promoter (e.g. the cauliflower mosaic virus 35S promoter) and the phenotypic expression observed. The ultimate goal of the reverse genetics approach is to determine gene function. Some reverse genetics strategies, such as targeted-induced local lesions in genomes (TILLING) [56] and RNA interference (RNAi) [57,58], could be used in comprehensive weed genomics projects. In TILLING, point mutations can be created in the weed genome and mutant candidate genes screened to identify potential weediness sequences with phenotypes. It has recently been suggested that 'ecotilling' could be used to study nucleotide polymorphisms in natural populations of plants [59]. Although the TILLING technique was first developed in *Arabidopsis*, it has been used successfully in many organisms. RNAi is a transgenic approach to induce double-stranded RNA to be formed for specific weediness gene candidates (once identified). The double-stranded RNA forms short interfering RNA signalling for degradation of specific mRNAs. Thus, weediness genes could be disrupted and observations made about whether the plant becomes less weedy. RNAi has not been performed in weeds.

Conclusions

As well as providing a basic understanding of the genetic basis of weediness, the development of weed genomics would also provide three predictable and useful outcomes. The first is the identification of genes that could improve crop yields. For example, several weeds have aluminum tolerance, a trait that is virtually absent from crops [60]. The second is to improve our understanding of the evolution of herbicide resistance and to aid in the identification of novel herbicide targets [4,61]. Currently, there is little (if any) solid predictive capability of why some weeds develop resistance and others do not (Box 2). Third, our understanding of weed biology would be exponentially expanded.

There will be hurdles to developing weed genomic models. Manipulated plant cultural conditions are needed to cycle weedy genotypes rapidly and to keep them small until they reproduce (not to be confused with genetic rapid cyclers that have lost weediness characteristics [62]). The manipulation of photoperiod would be one route towards rapid cycling. Self-incompatible, dioecious and polyploid species would also be problematic laboratory models because they impose constraints on facile manipulation, which drive up costs. Thus, weed genomics will not be as straightforward as *Arabidopsis* genomics, but the payoffs should ultimately outweigh any experimental inconveniences. The power of genomics is the integration of data [63]. Thus, we should learn much from the commonalities between taxa, collectively composed of genes, regulatory networks and other genomic factors responsible for weediness. These commonalities might be exploited as new strategies for sustainable weed management.

Perhaps the greatest impediment to acquiring genomic information about weeds and the development of models is inadequate funding. International consortia need to be formed to convince appropriate funding agencies that understanding weed genomes will not only improve agriculture, but also yield important information about

competition and plant biology. US$20 billion of damages in the USA alone should carry some weight.

Acknowledgements

We appreciate helpful conversations with Suzanne Warwick, Larry Steckel, Scott McElroy and Chris Main. The manuscript was greatly improved by helpful suggestions by Jonathan Gressel and two anonymous reviewers. Our interest in this topic has been spurred by ten years of funding from the USDA Biotechnology Risk Assessment Research Grants Program, for whose grants we are thankful. A symposium on 'Weed Genomics' will be held in conjunction with the Weed Science Society of American annual meeting (February 2005, Honolulu, HI, USA).

References

1 Bridges, D.C. (1994) Impact of weeds on human endeavors. *Weed Technol* 8, 392–395
2 Holm, L.G. *et al*. (1977) *World's Worst Weeds: Distribution and Biology*, University Press of Hawaii
3 Holm, L. *et al*. (1997) *World Weeds: Natural Histories and Distribution*, John Wiley & Sons
4 Gressel, J. (2002) *Molecular Biology of Weed Control*, Taylor & Francis
5 Ross, M.A. and Lembi, C.A. (1999) *Applied Weed Science*, Prentice Hall
6 Marshall, G. (2001) A perspective on molecular-based research: integration and utility in weed science. *Weed Sci.* 49, 273–275
7 Dyer, W.E. (1991) Application of molecular biology in weed science. *Weed Sci.* 39, 482–488
8 Paltridge, N.G. (2000) Applications for molecular biology in weed management. *Plant Prot. Quart.* 15, 50–56
9 Jasieniuk, M. and Maxwell, B.D. (2001) Plant diversity: new insights from molecular biology and genomics technologies. *Weed Sci.* 49, 257–265
10 Weller, S.C. *et al*. (2001) The effect of genomics on weed management in the 21st century. *Weed Sci.* 49, 282–289
11 Nissen, S.J. *et al*. (1995) DNA-based marker systems to determine genetic diversity of weedy species and their application to biocontrol. *Weed Sci.* 43, 504–513
12 O'Hanlon, P.C. *et al*. (2000) A review of new PCR-based genetic markers and their utility to weed ecology. *Weed Res.* 40, 239–254
13 Maloof, J.N. (2003) QTL for plant growth and morphology. *Curr. Opin. Plant Biol.* 6, 85–90
14 Foley, M.E. (2002) Weeds, seeds, and buds: opportunities and systems for dormancy investigations. *Weed Sci.* 50, 267–272
15 Chao, W.S. (2002) Contemporary methods to investigate seed and bud dormancy. *Weed Sci.* 50, 215–226
16 Coleman, R.K. *et al*. (2001) Identification of quantitative trait loci for traits conferring weed competitiveness in wheat (*Triticum aestivum* L.). *Aust. J. Agric. Res.* 52, 1235–1246
17 Moritz, D.M.L. and Kadereit, J.W. (2001) The genetics of evolutionary change in *Senecio vulgaris* L.: a QTL mapping approach. *Plant Biol.* 3, 544–552
18 Fennimore, S.A. *et al*. (1999) A genetic model and molecular markers for wild oat (*Avena fatua* L.) seed dormancy. *Theor. Appl. Genet.* 99, 711–718
19 Paterson, A.H. *et al*. (1995) The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) Pers. *Proc. Natl. Acad. Sci. U. S. A.* 92, 6127–6131
20 Hu, F.Y. *et al*. (2003) Convergent evolution of perenniality in rice and sorghum. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4050–4054
21 Horvath, D.P. and Anderson, J.V. (2002) A molecular approach to understanding root bud dormancy in leafy spurge. *Weed Sci.* 50, 227–231
22 Horvath, D.P. *et al*. (2003) Identification of genes induced in emerging tillers of wild oat (*Avena fatua*) using *Arabidopsis* microarrays. *Weed Sci.* 51, 503–508
23 Horvath, D.P. *et al*. (2003) *Arabidopsis* microarrays identify conserved and differentially expressed genes involved in short growth and development from distantly related plant species. *Plant J.* 34, 125–134
24 The *Arabidopsis* Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
25 Meinke, D.W. *et al*. (1998) *Arabidopsis thaliana*: a model plant for genome analysis. *Science* 662, 679–682
26 Yu, J. *et al*. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296, 79–92
27 Goff, S.A. *et al*. (2002) A draft sequences of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296, 92–100
28 Patzoldt, W.L. *et al*. (2002) Variable herbicide responses among Illinois waterhemp (*Amaranthus rudis* and *A. tuberculatus*) populations. *Crop Prot.* 21, 707–712
29 VanGessel, M.J. (2001) Glyphosate-resistant horseweed from Delaware. *Weed Sci.* 49, 703–705
30 Mueller, T.C. *et al*. (2003) Shikimate accumulates in both glyphosate-sensitive and glyphosate-resistant horseweed (*Conyza canadensis* L. Cronq). *J. Agric. Food Chem.* 51, 680–684
31 Halfhill, M.D. *et al*. (2002) *Bt*-transgenic oilseed rape hybridization with its weedy relative, *Brassica rapa*. *Environ. Biosafety Res* 1, 19–28
32 Stewart, C.N., Jr. *et al*. (2003) Transgene introgression from genetically modified crops to their wild relatives. *Nat. Rev. Genet.* 4, 806–817
33 Quiros, C.F. *et al*. (2001) *Arabidopsis* and *Brassica* comparative genomics: sequence, structure and gene content in the ABI1–Rps2–Ck1 chromosomal segment and related regions. *Genetics* 157, 1321–1330
34 Paterson, A.H. *et al*. (2001) *Brassica* genomics: a complement to, and early beneficiary of, the *Arabidopsis* sequence. *Genome Biol.* 2, 10111–10114
35 Vaughan, L.K. *et al*. (2001) Is all red rice found in commercial rice really *Oryza sativa? Weed Sci.* 49, 468–476
36 Bres-Patry, C. *et al*. (2001) Heredity and genetic mapping of domestication-related traits in a temperate *japonica* weedy rice. *Theor. Appl. Genet.* 102, 118–126
37 Rieseberg, L.H. *et al*. (2000) Hybridization, introgression, and linkage evolution. *Plant Mol. Biol.* 42, 205–224
38 Gu, X-Y. *et al*. (2004) Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (*Oryza sativa*). *Genetics* 166, 1503–1516
39 Pryer, K.M. *et al*. (2002) Deciding among green plants for whole genome studies. *Trends Plant Sci.* 7, 550–554
40 Green, E.D. (2001) Strategies for the systematic sequencing of complex genomes. *Nat. Rev. Genet.* 2, 573–583
41 Zemetra, R.S. *et al*. (1998) Potential for gene transfer between wheat (*Triticum aestivum*) and jointed goatgrass (*Aegilops cylindrica*). *Weed Sci.* 46, 313–317
42 Xiu, J.L. *et al*. (1997) Inheritance in synthetic hexaploid wheat 'RSP' of sprouting tolerance derived from *Aegilops tauschii* Cosson. *Euphytica* 95, 321–323
43 Fukao, T. *et al*. (2003) Differential gene expression of the alpha-chain of mitochondrial H^+-transporting ATP synthase between dormant and non-dormant seeds of paddy *Echinochloa* weeds. *Weed Biol. Manage.* 3, 15–20
44 Hayashi, H. *et al*. (1992) Activation of a plant gene by T-DNA tagging: auxin-independent growth *in vitro*. *Science* 258, 1350–1352
45 Krysan, P.J. *et al*. (1999) T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* 11, 2283–2290
46 Borsics, T. *et al*. (2002) Methods for genetic transformation of the parasitic weed dodder (*Cuscuta trifolii* Bab. et Gibs) and for PCR-based detection of early transformation events. *Plant Sci.* 162, 193–199
47 Pelletier, G. and Bechtold, N. (2003) *In planta* transformation. In *Transgenic Plants: Current Innovations and Future Trends* (Stewart, C.N., Jr., ed.), pp. 65–82, Horizon Scientific Press
48 Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743
49 Wang, W.C. *et al*. (2003) Development of a novel *Agrobacterium*-mediated transformation method to recover transgenic *Brassica napus* plants. *Plant Cell Rep.* 22, 274–281
50 Anderson, W.P. (1996) *Weed Science: Principles and Applications*, West Publishing Company

51 Fu, Y.B. and Ritland, K. (1994) Marker-based inferences about fecundity genes contributing to inbreeding depression in *Mimulus guttatus*. *Genome* 37, 1005–1010
52 Yanovsky, M.J. and Kay, S.A. (2002) Molecular basis of seasonal time measurement in *Arabidopsis*. *Nature* 419, 308–312
53 Suárez-López, P. (2002) *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* 410, 1116–1120
54 Yan, L. *et al*. (2004) The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science* 303, 1640–1644
55 Alonso-Blanco, C. *et al*. (2003) Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis thaliana*. *Genetics* 164, 711–729
56 McCallum, C. (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiol.* 123, 439–442
57 Fire, A. *et al*. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811
58 Waterhouse, P.M. and Helliwell, C.A. (2003) Exploring plant genomes by RNA-induced gene silencing. *Nat. Rev. Genet.* 4, 26–35
59 Comai, L. *et al*. (2004) Efficient discovery of DNA polymorphisms in natural populations by ecotilling. *Plant J.* 37, 778–786
60 Bilski, J.J. and Foy, C.D. (1988) Differential tolerances of weed species to aluminum manganese and salinity. *J. Plant Nutr.* 11, 93–106
61 Lein, W. *et al*. (2004) Target-based discovery of novel herbicides. *Curr. Opin. Plant Biol.* 7, 1–7
62 Williams, P.H. and Hill, C.B. (1986) Rapid-cycling populations of *Brassica*. *Science* 232, 1385–1389
63 Osterlund, M.T. and Paterson, A.H. (2002) Applied plant genomics: the secret is integration. *Curr. Opin. Plant Biol.* 5, 141–145
64 Warwick, S.I. *et al*. (2003) Hybridization between transgenic Brassica napus L. and its wild relatives: *Brassica rapa* L., *Raphanus raphanistrum* L., *Sinapis arvensis* L., and *Erucastrum gallicum* (Willd.) O.E. Schulz. *Theor. Appl. Genet.* 107, 528–539
65 Main, C.L. *et al*. (2004) Response of selected horseweed (*Conyza canadensis* (L.) Cronq.) populations to glyphosate. *J. Agric. Food Chem.* 52, 879–883
66 Feng, P.C.C. *et al*. Investigations into the mechanism of glyphosate resistance in horseweed (*Conyza canadensis*): retention, uptake, translocation and metabolism. Weed Sci. (in press)
67 Baker, H.G. (1965) Characteristics and modes of origin of weeds. In *The Genetics of Colonizing Species* (Baker, H.G. and Stebbins, G.L. eds), pp. 147–168, Academic Press
68 Zimdahl, R.L. (1993) *Fundamentals of Weed Science*, Academic Press
69 Knake, E.L. and Slife, F.W. (1962) Competition of *Setaria faberi* with corn and soybeans. *Weeds* 10, 26–29
70 Jeschke, M.R. *et al*. (2003) DNA content analysis of smooth pigweed (*Amaranthus hybridus*) and tall waterhemp (*A. tuberculatus*): implications for hybrid detection. *Weed Sci.* 51, 1–3
71 Federov, A.A. (1969) *Chromosome Numbers of Flowering Plants*, Academy of Sciences of USSR Komarov Botanical Institute
72 Galbraith, D.W. *et al*. (1983) Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* 220, 1049–1051