Summary

1. Herbivores exert significant selection on plants, and plants have evolved a variety of constitutive and inducible defences to resist and tolerate herbivory. Assessing the genetic mechanisms that influence defences against herbivores will deepen our understanding of the evolution of essential phenotypic traits.

2. Ecogenomics is a powerful interdisciplinary approach that can address fundamental questions about the ecology and evolutionary biology of species, such as: which evolutionary forces maintain variation within a population? and What is the genetic architecture of adaptation? This field seeks to identify gene regions that influence ecologically important traits, assess the fitness consequences under natural conditions of alleles at key quantitative trait loci (QTLs), and test how the abiotic and biotic environment affects gene expression.

3. Here, we review ecogenomics techniques and emphasize how this framework can address long-standing and emerging questions relating to anti-herbivore defences in plants. For example, ecogenomics tools can be used to investigate: inducible vs. constitutive defences; tradeoffs between resistance and tolerance; adaptation to the local herbivore community; selection on alleles that confer resistance and tolerance in natural populations; and whether different genes are activated in response to specialist vs. generalist herbivores and to different types of damage.

4. Ecogenomic studies can be conducted with model species, such as Arabidopsis, or their relatives, in which case myriad molecular tools are already available. Burgeoning sequence data will also facilitate ecogenomic studies of non-model species. Throughout this paper, we highlight approaches that are particularly suitable for ecological studies of non-model organisms, discuss the benefits and disadvantages of specific techniques and review bioinformatic tools for analysing data.

5. We focus on established and promising techniques, such as QTL mapping with pedigreed populations, genome wide association studies, transcription profiling strategies, population genomics and transgenic methodologies. Many of these techniques are complementary and can be used jointly to investigate the genetic architecture of defence traits and selection on alleles in nature.

Key-words: candidate gene, genome wide association studies, plant defence, population genomics, transgenics, quantitative trait loci

Introduction

Herbivores exert significant selection on plant populations, and plants have evolved a variety of constitutive and inducible defences to resist and tolerate herbivory (e.g. Shonle & Bergelson 2000; Juenger & Bergelson 2000; Tiffin 2000; Carmona, Lajunesse & Johnson 2010; Fornoni, 2011). Plant populations exhibit substantial phenotypic and genetic variation in chemical and physical defences, such as secondary metabolites, trichomes, lignin and C : N ratios, spines and thorns, capacity for regrowth and phenology (Kliebenstein, Figuth & Mitchell-Olds 2002a; Kroymann & Mitchell-Olds 2005; Agrawal & Fishbein 2006; Schranz et al. 2009). Phenotypic variation can be maintained through balancing, frequency-dependent, or divergent selection; alternatively, deleterious polymorphisms can be
produced via immigration of maladapted alleles, or by mutation–selection balance (Feder & Mitchell-Olds 2003; Mitchell-Olds, Willis & Goldstein 2007). Distinguishing among these possibilities will illuminate the relative roles of selection and genetic drift in evolution. An ecogenomics framework is well-suited for addressing emerging issues related to plant defences against herbivores.

Evolutionary and ecological functional genomics integrates across disciplines including evolutionary biology, ecology, genetics and physiology to investigate adaptive evolution and patterns of genetic and phenotypic variation within and among species (Feder & Mitchell-Olds 2003). This field seeks to identify gene regions associated with ecologically important traits, assess the fitness consequences of alleles at key quantitative trait loci (QTLs) under natural conditions and test how the abiotic and biotic environment influences gene expression. Ecogenomics techniques can be used to investigate questions that have long interested biologists, including: What is the relative importance of genetic drift and selection in population differentiation and ultimately speciation? What is the genetic architecture of adaptation (e.g. few genes with large effects, or many genes with small effects)? Do different taxa use similar genetic pathways to achieve convergent phenotypes? (Feder & Mitchell-Olds 2003; Mitchell-Olds & Schmitt 2006; Mitchell-Olds, Willis & Goldstein 2007; Stinchcombe & Hoekstra 2008; Agrawal 2011).

Ecogenomic studies can either focus on the evolution of a polymorphic trait (e.g. Schranz et al. 2009), or can use a transgenic approach to investigate gene function and ecological consequences (e.g. Baldwin et al. 2006). These methods allow researchers to address a number of key questions about plant defences, including issues pertaining to tolerance (maintenance of fitness after herbivory) of and resistance (deterrence of herbivory), such as: Do QTLs for tolerance and resistance co-localize? Are there molecular signatures of selection acting on alleles at these QTLs (e.g. De Meaux & Mitchell-Olds 2003)? Are tradeoffs between resistance and tolerance evident at the molecular level? In heterogeneous landscapes, are alleles for herbivore resistance and/or tolerance more prevalent in sites with higher herbivory pressure? Do similar pathways influence defence against herbivores and pathogens?

An ecogenomics approach can also investigate constitutive and inducible defences, thereby contributing to our general understanding of the functional basis of complex traits and addressing fundamental questions about the evolution of phenotypic plasticity (Holeski, Chase-Alone & Kelly 2010). Theory predicts that consistently high herbivore pressure favours constitutive defences, whereas low or variable herbivore pressure could favour inducible defences, if the costs of defence are high (e.g. Stamp 2003; Holeski, Chase-Alone & Kelly 2010; Karban 2011). Ecogenomics tools can be used to assess whether temporal and spatial variability in herbivory influences the type of defence that evolves, and can detect similarities and differences in the genetic architecture of inducible and constitutive defences. An ecogenomic approach can identify whether: similar genome regions influence both constitutive and inducible defences, genes associated with constitutive and inducible defences interact epistatically, and there is evidence for selection on these genes under natural conditions.

An ecogenomics framework can address a number of other issues, including: whether gene expression patterns differ as a function of the type of herbivore (e.g. specialist vs. generalist, choker vs. phloem feeder, Reymond et al. 2004; Thompson & Goggin 2006); what cues plants use to detect herbivory (saliva, leaf removal, etc.) (De Meaux & Mitchell-Olds 2003) and whether plants are capable of ‘eavesdropping’ on damaged neighbours via volatile organic compounds produced in response to herbivory (Baldwin et al. 2006; Karban 2011). Forward genetic techniques, such as QTL mapping, investigate the genetic basis of phenotypic variation, whereas reverse genetic techniques, such as gene silencing and other molecular approaches, attempt to identify the phenotypic effects of known genes. Both forward and reverse genetic techniques can be brought to bear on these ecological questions and elucidate the molecular underpinnings of plant defensive traits.

In model organisms and crop species, great strides have been made in understanding the evolution of plant defence, as well as gene expression changes in response to herbivory (e.g. Reymond et al. 2000; Schmidt et al. 2004; Thompson & Goggin 2006; McKay & Stinchcombe 2008). Molecular tools developed for model organisms can be applied in ecogenomics studies of closely related non-model species (Mitchell-Olds 2001); in this way, ecogenomic tools have illuminated the genetic basis of plant defensive traits in *Boechera, Brassica* and other members of the Brassicaceae, relatives of *Arabidopsis* (Broekgaarden et al. 2007; Kivimaki et al. 2007; Schranz et al. 2009). Furthermore, due to recent technological advances, ecogenomic tools are becoming cost effective and feasible for ecological studies of non-model systems that are only distantly related to model organisms. Many questions in the ecology and evolution of plant defence test historical hypotheses, such as how particular patterns evolved or why they exist. Such questions are best addressed in native species in relatively undisturbed environments, where ecological and genetic factors may be near equilibrium. In contrast, studies of introduced and weedy species elucidate ‘real-time’ ecological interactions, but may reflect transient population dynamics rather than long-term evolutionary processes.

In this review, we discuss established and emerging methodologies for exploring plant defence from an ecogenomics perspective and we highlight approaches that are suitable for different types of study species (Feder & Mitchell-Olds 2003). Specifically, we focus on QTL mapping with pedigreed and natural populations, transcription profiling, population genomics and transgenic approaches (summarized in Table 1).

**Quantitative trait loci**

Quantitative trait locus mapping is a fundamental tool for investigating the evolution of polymorphisms in natural populations and the genetic basis of traits and fitness
Table 1. Overview of ecogenomic techniques reviewed in this paper

<table>
<thead>
<tr>
<th>Ecogenomic tool</th>
<th>Specific methodologies</th>
<th>Ecological and evolutionary questions</th>
<th>Case studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative trait locus (QTL) mapping: Identify gene regions that underlie phenotypic traits. Cannot be applied to monomorphic populations</td>
<td>Family-based approach; Population-based approach</td>
<td>Do different gene regions influence phenotypic traits in contrasting environments? What is the genetic architecture of adaptation (e.g., one QTL with large effects, or many QTLs of minor effect)? Do QTLs for defensive traits co-localize with QTLs for fitness components? Is there evidence for selection on QTL alleles in their native environment? Do QTLs for defence traits show a fitness cost when herbivores are absent? Is there evidence for tradeoffs between herbivore resistance and tolerance? That is, do QTLs for resistance and tolerance co-localize and exhibit negative genetic correlations?</td>
<td>Kliebenstein et al. 2002b; Mauricio 2005; Chan, Rowe &amp; Kliebenstein 2010; Schranz et al. 2009; Holeski, Chase-Alone &amp; Kelly 2010</td>
</tr>
<tr>
<td>Transcription profiling: Quantify gene expression differences among environments, treatments, life history stages or tissue types. Can be applied to monomorphic populations</td>
<td>Microarrays, SAGE, Next Generation Sequencing</td>
<td>Which genes are activated by herbivory? Do specialist and generalist herbivores induce different gene expression patterns? Do locally adapted populations differ in gene expression profiles? Are interpopulational differences in herbivore species composition or abundance reflected in the expression of plant defensive genes? Does selection favour defensive genes in the presence of herbivores?</td>
<td>Reymond et al. 2000; Voelckel &amp; Baldwin 2004; Wang, He &amp; He 2005; Thompson &amp; Goggin 2006; Broekgaarden et al., 2008, Broekgaarden et al. 2010</td>
</tr>
<tr>
<td>Population genomics: Test whether candidate loci differ from genome-wide demographic patterns to detect whether these loci are under selection</td>
<td>Genome-wide genotyping or sequencing</td>
<td>Is there a molecular signature of natural selection in certain gene regions, e.g. significant defence QTLs or candidate genes? Are candidate genes significantly associated with ecologically relevant traits? Do candidate genes exhibit different evolutionary histories than other genes? Is allelic variation at candidate loci associated with herbivore abundance, species composition or type (specialist vs. generalist)?</td>
<td>Moeller &amp; Tiffin 2005; Caldwell &amp; Michelmore 2009 Tiffin, Hacker &amp; Gaut 2004; Kivimaki et al. 2007</td>
</tr>
<tr>
<td>Candidate genes: Analyse association between candidate genes and phenotypic traits</td>
<td>Heterologous expression, gene silencing and overexpression</td>
<td>What is the phenotypic effect of a specific gene? (e.g., useful for confirmation of gene function after identification of genes in QTL mapping)</td>
<td>Krügel et al. 2002; Halitschke &amp; Baldwin 2003; Tian et al. 2003; Li, Brader &amp; Palva 2004; Kessler, Halitschke &amp; Baldwin 2004; Steppuhn et al. 2004</td>
</tr>
</tbody>
</table>

SAGE, Serial analysis of gene expression.
components. This methodology requires phenotyping and genotyping individuals to identify chromosomal regions that underlie differentiation in the trait of interest. Furthermore, genotype (i.e. QTL) by environment interactions can illuminate the genetic basis of local adaptation, and can facilitate analysis of phenotypically plastic traits such as inducible plant defences (Mitchell-Olds, Willis & Goldstein 2007). QTL mapping can be conducted using either pedigreed experimental populations (family-based approach), or unrelated individuals from natural populations (population-based approach); below, we review these approaches.

**Family-based approach**

Quantitative trait loci mapping with experimental pedigrees is a powerful way to detect the position and effect of genomic regions that influence plant defensive traits. In inbreeding-tolerant species, crosses between two divergent parents can produce a hybrid F1 generation, which can be self-pollinated to produce a heterozygous F2 generation, and recombinant inbred lines (RILs) after 6–8 generations (Collard et al. 2005). RILs provide ‘immortal’ genotypes that can be scored once for molecular markers, and subsequently phenotyped for many traits in many environments by a broad research community. By subsequent crossing, self-pollination, genotyping and selection of progeny, near-isogenic lines (NILs) can be generated. Such NILs segregate for two contrasting alleles in a QTL region, while being isogenic (identical) elsewhere in the genome. In the pre-genomics era, the segregating QTL region typically confounded hundreds of loci with the quantitative trait gene of interest. However, current technology enables NILs that differ by only one or a few genes (Fridman et al. 2004). NILs are time- and labour-intensive to generate (Tuinstra, Ejeta & Goldsborough 1997). Instead, HIFs (heterogeneous inbred families) that segregate only near the QTL of interest can be identified from partially inbred lines to produce NILs for fine-mapping (Tuinstra, Ejeta & Goldsborough 1997). Essentially, RILs that are heterozygous for the QTL of interest, but homozygous at other loci, are propagated to produce families (HIFs) that differ only in genotype at that QTL. After multiple generations of selling, the genome is expected to be mostly homozygous (expected homozygosity in F6 RILs: 96.9% homozygous), which could reduce the possibility of encountering families heterozygous at the QTL of interest. However, if there are sufficient numbers of families in a study, then it may be possible to identify several families heterozygous at that QTL.

Mapping QTLs using inbred lines is ideal for investigating the genetic architecture of anti-herbivore defences when these traits have a strong genetic component and vary either within or among populations. Indeed, the family-based approach has been used in model and non-model systems to investigate anti-herbivore chemical and mechanical defences (e.g. Kroymann et al. 2003; Freeman et al. 2008; Lou et al. 2008) resistance to specialist and generalist herbivores (e.g. Lambrix et al. 2001; Kliebenstein et al. 2002b), tolerance of mammalian herbivores (e.g. Weinig, Stinchcombe & Schmitt 2003), co-localization of fitness and resistance QTLs (e.g. Weinig, Stinchcombe & Schmitt 2003), the genetic architecture of both constitutive and inducible defences (Holeski, Chase-Alone & Kelly 2010), and the influence of ontogeny on defensive traits (Mauricio 2005). For example, Kliebenstein et al. (2002b) discovered that Arabidopsis QTLs for glucosinolate production co-localized with those for resistance to a generalist, but not a specialist herbivore; this pattern is concordant with theory and suggests that chemical defences may be more effective against generalists than specialists (but see Carmona, Lajunesse & Johnson 2011).

Data analysis involves the creation of linkage maps with markers sufficiently closely spaced to facilitate QTL detection, e.g. 10 cM spacing generally gives good results (1 cM distance between markers on a linkage map translates into 1% recombination frequency) (Erickson et al. 2004; Collard et al. 2005), followed by QTL mapping. A preliminary round of QTL analysis can be conducted to determine whether additional markers should be scored to increase marker density near potentially significant QTLs. A variety of markers can be used, including AFLPs, RFLPs, SNPs, RAD markers and microsatellites, as long as the parental lines carry different alleles at each locus (e.g. see Erickson et al. 2004; Collard et al. 2005; Hohenlohe et al. 2010; for a discussion of the advantages and disadvantages of these markers). AFLPs and RAD markers can aid the initial construction of linkage maps in non-model species because they can be screened rapidly and inexpensively (Erickson et al. 2004; Collard et al. 2005; Herrera & Bazaga 2009; Hohenlohe et al. 2010). The detection and scoring of markers in non-model species will become easier in the near future as sequencing costs decrease.

Quantitative trait locus mapping is used to detect individual QTLs, identify epistatic interactions among loci (Kao, Zeng & Teasdale 1999), and test QTL × environment interactions. Significant QTL × E interactions indicate different loci are activated in distinct environments and/or gene expression varies with environment. Due to the inducible nature of many plant defences, it is important to test whether QTLs are constant across environments and/or herbivory treatments. It is not sufficient simply to compare QTLs mapped in different environments or under different treatments, because lack of concordance may be due to limited statistical power rather than genetic differences between environments. Explicit, robust statistical analyses of QTL × E interactions are available in free computer programs, such as qtl Cartographer (Basten, Weir & Zeng 2004), and the r/qtl package in the r statistical environment (Broman & Sen 2009). By mapping QTLs in the presence and absence of herbivores, and testing for QTL × E interactions, researchers can identify genome regions that influence induced defences.

**Population-based approach (Linkage disequilibrium mapping)**

Studies of *Populus*, *Eucalyptus*, *Pseudotsuga* and other trees illustrate that the family-based approach to QTL mapping can be implemented for some long-lived species (e.g. Wu
105 SNPs), which may not be available for studies of plant defense until next generation technologies enable rapid, inexpensive sequencing. An additional method has been used to identify QTLs in human populations, but has been underutilized in evolutionary studies (Mackay 2001; Erickson et al. 2004). The sib-pair method exploits the relationship between siblings to model the differences in phenotypes as a function of the loci for which a sib-pair has alleles identical by descent (Driigalenko 1998). Similar to GWAS, the sib-pair method can be used in outbred populations; however, it uses trait variation among siblings to map QTLs and therefore family relationships must be known. This method requires large numbers (hundreds to thousands) of sampled families to map QTLs reliably, and is unlikely to detect small effect QTLs (Fulker & Cherny 1996; Iliadou et al. 2007; Kleensang et al. 2010); nonetheless, it can be used to identify gene regions with major effects on phenotypes, and test for genotype by environment interactions (Erickson et al. 2004), which are especially important when studying inducible defences. Indeed, this method could be especially useful for plant populations, where there are many half-sib families (Erickson et al. 2004).

Quantitative trait locus mapping with pedigreed families and outbred populations are complementary approaches that can reveal genome regions that have large and small effects on phenotypes, and that are relevant to the evolution of variation in natural populations (Chan, Rowe & Kliebenstein 2010). Once a significant QTL is mapped, researchers can conduct additional studies to verify the QTL, isolate and clone the gene, and investigate the molecular biology and biochemistry of causal genes (Mackay 2001). However, hundreds of genes can be contained within the confidence limits of significant QTLs, and isolating the gene or genes that regulate phenotypic expression is not a trivial issue (Salvi & Tuberosa 2005). Limited genetic and sequence information will constrain functional studies in non-models systems and may prevent identification of the underlying gene(s). Nonetheless, studies can still address fundamental questions in ecology, evolution and plant–herbivore interactions. For example, field experiments using families that segregate for significant QTL alleles can assess the pleiotropic effects of these QTLs on fitness components, and investigate selection on QTLs in the presence and absence of herbivores to evaluate the costs of defence. Such studies can quantify tradeoffs between tolerance and resistance, e.g. by determining whether QTLs for herbivore resistance and tolerance co-localize and exhibit negative genetic correlations. A QTL approach could help determine whether negative genetic correlations would constrain the evolution of tolerance and resistance traits (e.g. Gardner & Latta 2007). These experiments will elucidate which evolutionary forces maintain phenotypic variation in populations, especially when lines are reciprocally transplanted into the parental environments; however, such studies have only rarely been conducted (Mitchell-Olds, Willis & Goldstein 2007).

**Transcription profiling**

Many plant species exhibit phenotypic plasticity in defensive traits in response to herbivory (e.g. Karban & Baldwin 1997). The genetic architecture of plasticity and induced defences can be investigated using transcription profiling techniques (Holeski, Chase-Alone & Kelly 2010). Transcription profiling quantifies expression of many genes in response to specific environments and stresses (e.g. herbivory), at different ontogenetic stages, or in different types of tissue, and can illuminate the genetic responses to local selection pressures (e.g. local herbivore species composition or abundance).
expression differences between undamaged and damaged plants to identify genes activated by herbivory (Broekgaarden et al. 2007), compared gene expression patterns between different cultivars exposed to herbivory in the field (Broekgaarden et al. 2010), and demonstrated that specialist and generalist herbivores induce different gene expression (Voelckel & Baldwin 2004). These methods are increasingly cost-effective and feasible for ecological studies of non-model species.

Expression microarrays are two dimensional arrays of genes used to quantify the expression of many genes simultaneously; this approach has been widely used for transcription profiling. Microarrays use several technologies to localize specific DNA sequences (probes) at known positions on a two-dimensional surface (Leung & Cavaliere 2003). mRNA extracted from a sample is converted to cDNA, labelled with a fluorescent dye, and hybridized to the probes on the microarray, leading to localized fluorescent signals indicative of expression levels for each gene. This approach allows researchers to detect relative expression differences treatments and have been up- or down-regulated compared with a control (Leung & Cavaliere 2003). Arrays are available commercially for model organisms, but also can be customized for non-model species. Brazma et al. (2001) and Leung & Cavaliere (2003) discuss general guidelines for experimental design of microarray studies and detail the information that needs to be reported so that results are reproducible and understandable across labs.

Due to the enormous amount of data produced by microarray studies and the possibility of contamination and low-quality results, microarray experiments need to be planned well, have adequate biological replication, and statistical analyses need to be conducted carefully (Brazma et al. 2001; Chuaqui et al. 2002; Jain et al. 2003; Leung & Cavaliere 2003; Wouters et al. 2003; Wei, Li & Bumgarner 2004; Vardhanabhuti et al. 2006). Once a high-quality image is produced and analysed by imaging software, data can be plotted and analysed in several platforms, including the statistical microarray analysis (SMA) package in R and the Statistical Analysis of Microarray (SAM) program (Storey & Tibshirani 2003). Care must be taken to account for multiple statistical tests, quantify false-discovery rates and reduce type I and II errors. Genes that exhibit differential expression should be evaluated and verified with techniques such RT-PCR (e.g. Chuaqui et al. 2002).

Microarray studies require known genes relevant to a specific treatment or environmental comparison. Currently, microarray studies are practical for model organisms and their close relatives. This technique is becoming feasible for studies of non-model species as genomic libraries of expressed genes are developed via next generation sequencing (libraries of expressed sequence tags, EST) (e.g. Tittiger 2004). Microarrays can identify genes that differ in expression (i.e. mRNA abundance) between groups, but cannot detect differences in alleles that have similar mRNA expression levels (Erckson et al. 2004).

Microarray studies have been used to characterize induced defences in plant populations (Ralph et al. 2006; Thompson & Goggin 2006; Broekgaarden et al. 2007; Philippe et al. 2009), compare transcriptional responses to specialist and generalist herbivores (Reymond et al. 2004; Voelckel & Baldwin 2004), demonstrate coordinated defences in response to pathogens and herbivores (Schenk et al. 2000), and detail differences in gene expression between species of plants in response to the same herbivore (Schmidt et al. 2005). For example, Broekgaarden et al. (2010) employed microarrays to study defences against herbivory in Brassica oleracea cultivars in the field. Interestingly, the cultivar with lower expression of defensive genes had significantly greater herbivore load and species richness than the cultivar with high defensive gene expression (Broekgaarden et al. 2010), thus demonstrating community-level effects of gene expression differences in plants. Microarray studies have also demonstrated that herbivory and mechanical damage elicit different gene expression profiles, suggesting that plants respond to compounds released while herbivores consume plant tissue (reviewed in Korth 2003). Thus, mechanical damage does not entirely replicate herbivory and experiments should be designed accordingly.

Recent ‘genetical genomics’ approaches use gene expression data as quantitative traits to map ‘eQTLs’, the QTLs that control expression level of a particular gene (Erckson et al. 2004; Dallas et al. 2005; Kammenga et al. 2007; Shiu & Borevitz 2008; Kliebenstein 2009). Kliebenstein et al. (2006) mapped QTLs that influence the expression of glucosinolate genes in Arabidopsis with this approach.

Serial analysis of gene expression (SAGE) is a sequencing-based technique that permits rapid analysis of thousands of gene transcripts (Velculescu et al. 1995; Meyers et al. 2004; Thomas & Klapner 2004; Tittiger 2004; Wang 2007). SAGE can be applied to ecological studies of non-model organisms because prior knowledge of gene sequences is not necessary (Wang 2007). This technique involves isolation of mRNA, reverse transcription to cDNA, extraction of a small tag segment (c. 15 based pairs) for gene identification, ligation of all extracted tags into long chains (concatemers) for sequencing and quantification of gene expression based on the number of times a tag is sequenced (Velculescu et al. 1995). SAGE is more expensive than DNA microarrays, but it allows absolute (rather than relative) quantification of gene expression (Wang 2007). A recent development called SuperSAGE allows tags to be longer (25 bp), which improves the specificity of the tags and makes this method more suitable for non-model species (Matsumura et al. 2003) and SuperSAGE arrays integrate SuperSAGE and DNA microarray technologies to quantify gene expression (Matsumura et al. 2006). SAGE can be used to address key evolutionary questions,
including: Does expression of defence genes in plants differ as a function of herbivore diversity, species identity, or abundance? or Are different genes up- or down-regulated in response to herbivory by specialists vs. generalists? SAGE methodologies have been applied in studies of model and non-model organisms, including investigations of plant disease resistance (Matsumura et al. 2003; Restrepo et al. 2005; Hamada et al. 2008), developmental changes in gene expression (Obermeier et al. 2009) drought stress (Molina et al. 2008), and anti-herbivore defences (Gilaradoni et al. 2010). Wang (2007) comprehensively reviews the basic features of SAGE.

Other technologies have been applied for analyses of gene expression differences, and have played a supportive role in development of expression arrays for ecological model organisms. cDNA-AFLPs detect differences in gene expression through comparisons of cDNA band intensities on a gel (Ouborg & Vriezen 2007). This technique has been useful when sequence data are not available for a taxon and it has been successful in identifying (novel) genes associated with ecologically relevant traits (Donson et al. 2002; Volkmath et al. 2003; Bae et al. 2006; Knight et al. 2006; Ouborg & Vriezen 2007). In addition, suppression-subtractive hybridization (SSH) allows the detection of genes that are expressed differentially between several conditions or tissues (Diatchenko et al. 1996). This method can be applied to non-model organisms, and has been used to study gene expression in response to herbivory (Lawrence & Novak 2004; Wang, He & He 2005). Dedicated microarrays have been constructed with sequences discovered via cDNA-AFLP or SSH analyses.

Next-generation sequencing, currently available on several platforms (reviewed in Hudson 2008), offers a cost-effective, high throughput, alternative to traditional Sanger sequencing (Morozova & Marra 2008; Wang, Gerstein & Snyder 2009). These platforms generate deep coverage of shorter reads than Sanger sequencing, which can be used to identify polymorphisms (e.g. SNPs), determine the structure of the DNA template (e.g. the existence of exons vs. introns), and measure the transcriptome (Morozova & Marra 2008; Mackay, Stone & Ayroles 2009). Next generation sequencing quantifies expression levels of thousands of loci by counting the number of transcripts encountered for each gene, in a process known as transcriptome sequencing (Hudson 2008), ‘RNA-Seq’ (Wang, Gerstein & Snyder 2009) or the sequence census approach (Morozova & Marra 2008). This approach is functionally analogous to SAGE, but has the advantage of being capable of detecting rare transcripts (Morozova & Marra 2008). It has enormous potential for studies of non-model organisms because transcript abundance can be quantified in the absence of genomic sequence data (e.g. Vera et al. 2008; Ekblom et al. 2010; O’Neil et al. 2010). Sequencing of the transcriptome has already been applied to analysis of secondary metabolism in plant species (Sun et al. 2010). As sequencing costs decline, it is likely that this approach will replace competing technologies for transcription profiling. Bioinformatics tools to analyse vast amounts of sequence data are currently being developed (e.g. Mortazavi et al. 2008; Zhou et al. 2009), and include packages in R, such as GenomeGraphs and Genominator (http://www.bioconductor.org/help/course-materials/2008/SeattleNov08/RNASeq/).

CHALLENGES

Historically, high costs of transcription profiling have prevented the levels of experimental replication which are desirable for complex, multivariate traits in ecology. High throughput sequencing technologies will remove this limitation in the near future, enabling analyses of plant responses to herbivory under field conditions. A more difficult problem is to disentangle expression differences among gene family members, which may have nearly identical coding sequences, but divergent expression due to regulatory changes. Although this issue can be resolved in model organisms with high-quality reference sequences, it remains a problem in wild species, especially in polyploid organisms. Finally, transcription profiling data are simply multivariate descriptions of gene expression. The challenge remains to apply these methods to hypothesis testing in ecology and evolution, although in theory, transcription profiling data should be useful for studying issues such as induced defences and adaptation to local herbivore communities.

Population genomics

A population genomics approach can be taken to discover genome regions that might be evolving under the influence of natural selection. Evolutionary forces like genetic drift and gene flow influence all regions of the genome in a similar fashion; natural selection, in contrast, does not act uniformly across the genome (Moeller & Tiffin 2005; Stinchcombe & Hoekstra 2008; Hohenlohe et al. 2010). Rather, selection may reduce genetic variation within a population, but increase it among populations occupying distinct habitats subject to different local selection pressures. Population genomics identifies loci that differ from typical genome-wide demographic patterns in population divergence (e.g. $F_{ST}$ values), nucleotide diversity, or other parameters (Moeller & Tiffin 2005; Stinchcombe & Hoekstra 2008; Hohenlohe et al. 2010). Presumably such outliers could be linked to loci under selection. Population genomic studies require large numbers of loci to be genotyped or sequenced, which will shortly become feasible even in non-model systems due to increasing availability of sequence data (Stinchcombe & Hoekstra 2008; Hohenlohe et al. 2010; Turner et al. 2010).

The population genomics approach, however, does not directly inform studies of phenotypic evolution because it screens the genome in the absence of trait data. Nevertheless, this approach could be used in conjunction with QTL mapping, candidate genes and other quantitative genetic approaches (Stinchcombe & Hoekstra 2008; Caldwell & Michelmore 2009). For example, in a study of plant-
pathogen interactions, Moeller & Tiffin (2005) tested for non-neutral evolution in candidate defence genes of two species of Zea by comparing genetic diversity and evolutionary history of these candidate genes with genes not implicated in defence. Likewise, Caldwell & Michelmore (2009) found different evolutionary patterns in the genes encoding defence signaling vs. pathogen recognition proteins in Arabidopsis. Both studies detected evidence for selection at some of the candidate defence loci studied (Moeller & Tiffin 2005; Caldwell & Michelmore 2009). Similar studies in plant-herbivore systems could elucidate whether plant defensive QTLs or candidate genes show molecular signatures of selection, and distinguish between coevolutionary arms races vs. selective maintenance of functional polymorphisms (Bakker et al. 2006).

Candidate gene approaches

Large numbers of candidate genes for ecologically relevant traits have been identified through extensive research with model organisms, such as Arabidopsis genes influencing trichome development (Payne, Zhang & Lloyd 2000; Stracke, Werber & Weisshaar 2001), production of volatile organic compounds (Van Poecke, Posthumus & Dicke 2001), glucosinolates (Halkier & Gershenzon 2006) and phytoalexins (Bottcher et al. 2009). Synteny (i.e. gene colocalization on homologous chromosomes in related species) can facilitate investigations of the genetic basis of phenotypic traits in non-model species and allow researchers to analyse the association between phenotypic variation and allelic variation at a candidate gene (Kivimaki et al. 2007), when species-specific versions of a gene can be detected using degenerate PCR primers (Ouborg & Vriezen 2007). For example, in a study of Arabidopsis lyrata, Kivimaki et al. (2007) found that leaf trichomes reduced herbivory in natural populations; they then assessed the association between trichome production and sequence variation at a gene homologous to the GLABROUS1 (GL1) gene, which influences trichome formation in A. thaliana. In an analysis controlling for population structure (Pritchard et al. 2000), Kivimaki et al. (2007) discovered that mutations in the regulatory region of GL1 influenced the production of glabrous (hair-less) leaves.

The candidate gene approach is especially illuminating when combined with other methods, including QTL mapping, population genomics and transcription profiling (Tiffin, Hacker & Gaut 2004). Candidate genes can be identified within the confidence intervals of significant QTLs and can assist in the eventual determination of the genes underlying traits (e.g. Faris et al. 1999). Population genomic studies can reveal whether candidate genes exhibit different evolutionary histories than putatively neutral loci (e.g. Tiffin, Hacker & Gaut 2004) and can assess the relationship between allelic variation at a candidate gene and latitudinal or other environmental variation (Storz 2005; Balasubramanian et al. 2006). Similarly, plant defence studies could determine whether allelic variation at candidate defence genes is associated with local herbivore abundance and/or species composition (e.g. different genes or patterns of gene expression might be activated by specialist vs. generalist herbivores). The candidate gene approach may be less informative for non-model organisms that are only distantly related to model species.

Transgenic approaches

Transgenic experiments are an important tool for targeted manipulation of traits and gene expression in ecological genomics. These approaches may involve insertion of a gene into another species for functional analysis, overexpressing or silencing endogenous genes, or complementation of a mutant allele by transformation (Rasmann & Agrawal 2009). Most of these techniques are only available for model organisms, but they could possibly be applied to non-model species if the sequence of candidate genes is known. These experiments are influenced by several sources of variation which must be controlled by experimental design, replication and statistical analysis. In most plant systems, transgene expression is influenced by chromosomal insertion site ‘position effects’ (Gelvin 2003; Butaye et al. 2005) and by ‘somaclonal variation’ i.e. the epigenetic changes in gene expression due to tissue culture (Kaeppler, Kaeppler & Rhee 2000). Such variation among transgenic lines is reportedly absent in Nicotiana attenuata (Schwachtje, Kutschbach & Baldwin 2008), although position effects and somaclonal variants are well known from other studies in tobacco (Mannerlof & Tenning 1997; Nikola et al. 1998). Appropriate controls, biological replication and clear description of statistical analyses remain essential. Transgenic approaches will be difficult for quantitative traits that are controlled by numerous genes that interact epistatically. Finally, because of the risk that foreign genes could escape into natural populations, experiments are usually terminated before reproduction, thus complicating estimates of plant fitness. The potential for this approach is illustrated by a landmark study (Tian et al. 2003), where a large, replicated transgenic field experiment showed a significant fitness cost of plant resistance, despite the existence of substantial position effects.

Heterologous expression involves experimentally transferring cDNA into a species that does not normally express a given gene, via cloning into a vector, transforming into a novel system and finally phenotyping to verify gene function (Xu et al. 2004; Schnee et al. 2006). This approach is a powerful way to test the phenotypic effects of a candidate gene, and has been used to study defensive traits in the Brassicaceae and Solanaceae (Schnee et al. 2006; Koller et al. 2008; Rasmann & Agrawal 2009). For example, Schnee et al. (2006) heterologously expressed a maize gene in Escherichia coli to characterize its function in the production of volatile organic compounds, and then transformed it into A. thaliana for bioassays with the parasitoid of maize herbivores. Female parasitoids (Cotesia margini-
ventris) were attracted to the maize volatiles produced by transgenic Arabidopsis, thereby revealing the key role of a single gene in the indirect defence of maize against its natural enemies (Schnee et al. 2006).

GENE SILENCING AND OVEREXPRESSION STUDIES

Candidate genes can also be silenced or overexpressed via stable transformation to investigate their effects on plant defence (Krügel et al. 2002; Halitschke & Baldwin 2003; Kessler, Halitschke & Baldwin 2004; Li, Brader & Palva 2004; Steppuhn et al. 2004). For example, gene silencing studies have shown reduced resistance to a specialist caterpillar due to decreased production of nicotine and trypsin protease inhibitors and diminished jasmonic acid signaling in Nicotiana attenuata (Halitschke & Baldwin 2003; Steppuhn et al. 2004; Zavala et al. 2004); gene silencing has also uncovered previously unknown interactions between trypsin protease inhibitors and nicotine in N. attenuata defence against a generalist herbivore (Steppuhn & Baldwin 2007). Similarly, an overexpression study demonstrated the importance of the jasmonic acid signaling pathway in defence against herbivores for Solanum lycopersicum (Chen et al. 2005). Microarray studies can be conducted in concert with transgenic experiments to detect other genes that differ in expression between wild type and transgenic plants (Halitschke & Baldwin 2003).

POST-TRANSCRIPTIONAL GENE SILENCING

Virus-induced gene silencing (VIGS) transiently suppresses gene function by exploiting the antiviral defences of plants (Benedito et al. 2004; Burch-Smith et al. 2004; Gould & Kramer 2007). In this approach, a candidate gene is cloned into a virus, and seedlings are subsequently infected with that virus (Burch-Smith et al. 2004). Antiviral defences attack not only the virus, but also degrade the mRNA produced by the candidate gene, which results in ephemeral gene silencing (Burch-Smith et al. 2004). Post-transcriptional gene silencing through VIGS is an efficient method of manipulating phenotypes for genomic studies, especially if stable transformation is unfeasible (Benedito et al. 2004), but it has several disadvantages, including: gene silencing is transient, the plant species must be susceptible to the virus, and it may not be suitable for field studies. VIGS studies of plant defence would need to be timed perfectly so that a candidate defence gene is silenced when treatments are imposed and for the duration of the experiment.

This emerging technique was used in a recent study of oviposition behaviour of the herbivore Pieris rapae on three species of plants (A. thaliana, Brassica nigra and Nicotiana benthamiana) (Zheng et al. 2010) and to silence genes encoding anti-herbivore defences in N. attenuata (Mitra et al. 2008). VIGS has also been used to characterize gene function in numerous other ecologically relevant situations, including under water stress and in disease resistance (Purkayastha & Dasgupta 2009).

CHALLENGES

To reveal the evolutionary significance of genes, transgenic approaches need to either: (i) manipulate existing alleles at polymorphic loci (e.g. Tian et al. 2003), or; (ii) recreate both ancestral and derived genotypes (Dean & Thornton 2007). The silencing or overexpression of only one gene in a complex biosynthetic pathway is unlikely to duplicate the ancestor, or reflect the evolutionary history of a lineage. Furthermore, reverse genetic studies could inadvertently silence paralogous genes whose functions have diverged from the gene of interest. This problem must be tested explicitly.

Future directions and conclusions

The approaches we have discussed are complementary. For example, QTL studies can suggest which genes to analyse on a microarray (Erickson et al. 2004). In conjunction with these approaches, researchers can capitalize on the existence of naturally occurring mutants to investigate the genetic architecture of plant defences. When embarking upon a new ecogenomics study, it may be useful to select a focal species that is closely related to a model organism, so that sequence data and genomic resources are readily available. Nevertheless, next generation sequencing platforms will vastly improve our ability to collect sequence data from non-model organisms, which, in turn, will enable more thorough investigations of evolution in natural populations.

Next generation sequencing will also facilitate comparative genomic studies (e.g. Barakat et al. 2009). For example, phylogenetic studies and expression profiling can reveal the evolutionary history of genes encoding secondary metabolites, quantify duplication events and address changes in function of duplicates (e.g. Barakat et al. 2009). Furthermore, for species whose genomes have been fully sequenced, whole genome resequencing (sequencing multiple individuals from a population) and resequencing at specific loci can advance our understanding of population structure, genetic variation and selection, and can facilitate fine mapping in QTL studies (Hudson 2008).

Rapid technological advances in genomics methods and bioinformatic tools make interdisciplinary collaborations crucial for successful implementation of ecogenomics studies. Additionally, ecologists and evolutionary biologists should be mindful of genomic approaches that are used in human populations, because humans share several characteristics with non-model species in ecological studies, including long generation times, complex demographic history and population structure, and inability to produce experimental inbred populations. In this review, we have focused on plant anti-herbivore defensive traits. Clearly, ecogenomics studies focusing on herbivores (e.g. Tittiger 2004) will enhance our understanding of coevolution of plants and their natural enemies (Bergelson, Dwyer & Emerson 2001). Furthermore, research on plant–pathogen interactions should inform ecogenomic studies of plant defences (e.g. Matsumura et al. 2003; Moeller & Tiffin 2005; Barakat...
et al. 2009; Caldwell & Michelmore 2009). An ecogenomics approach is essential to resolving long-standing questions in evolutionary biology. Ecogenomic tools can illuminate the evolutionary processes and genetic mechanisms that influence phenotypic variation, elucidate the molecular basis of phenotypic plasticity (genotype-by-environment interactions), and reveal the genomic signature of selection and the genetic architecture of adaptation in temporally and spatially heterogeneous landscapes.

Acknowledgements

The authors were supported by grants from NIH (R01 GM086496 to TMO) and NSF (EF-0723447 to TMO). We thank R. Colautti, T. Pendersgast and C. Rushworth for helpful discussions and comments. Suggestions from M. Johnson and three anonymous reviewers increased the quality of this manuscript.

References


