MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 2226–2240 doi: 10.1111/mec.13026

INVASION GENETICS: THE BAKER AND STEBBINS LEGACY

Comparative genomics in the Asteraceae reveals little evidence for parallel evolutionary change in invasive taxa

KATHRYN A. HODGINS,* DAN G. BOCK,† MIN A. HAHN,† SYLVIA M. HEREDIA,† KATHRYN G. TURNER† and LOREN H. RIESEBERG†

School of Biological Sciences, Monash University, Clayton, Vic. 3800, Australia, †Department of Botany, University of British Columbia, 1316-6270 University Blvd., Vancouver, British Columbia V6T 1Z4, Canada

Abstract

Asteraceae, the largest family of flowering plants, has given rise to many notorious invasive species. Using publicly available transcriptome assemblies from 35 Asteraceae, including six major invasive species, we examined evidence for micro- and macro-evolutionary genomic changes associated with invasion. To detect episodes of positive selection repeated across multiple introductions, we conducted comparisons between native and introduced genotypes from six focal species and identified genes with elevated rates of amino acid change (dN/dS). We then looked for evidence of positive selection at a broader phylogenetic scale across all taxa. As invasive species may experience founder events during colonization and spread, we also looked for evidence of increased genetic load in introduced genotypes. We rarely found evidence for parallel changes in orthologous genes in the intraspecific comparisons, but in some cases we identified changes in members of the same gene family. Using among-species comparisons, we detected positive selection in 0.003–0.69% and 2.4–7.8% of the genes using site and stochastic branch-site models, respectively. These genes had diverse putative functions, including defence response, stress response and herbicide resistance, although there was no clear pattern in the GO terms. There was no indication that introduced genotypes have a higher proportion of deleterious alleles than native genotypes in the six focal species, suggesting multiple introductions and admixture mitigated the impact of drift. Our findings provide little evidence for common genomic responses in invasive taxa of the Asteraceae and hence suggest that multiple evolutionary pathways may lead to adaptation during introduction and spread in these species.

Keywords: Asteraceae, genetic load, genomics, invasive species, positive selection, transcriptome

Received 1 October 2014; revision received 23 November 2014; accepted 26 November 2014

Introduction

Anthropogenic changes have transformed the global landscape and, although many species are suffering from habitat loss and extinction as a result, invasive species have thrived. Invasive species, those that spread outside their natural range and proliferate (Gray & Mack 1986), can have negative impacts on the economy and the environment, which provides considerable incentive to understand the factors that contribute to their success (Stewart *et al.* 2009). However, the study of invasive species offers opportunities beyond the applied realm. Several decades ago, Baker (1974) proposed that weedy and invasive species were excellent subjects of evolutionary study because of their abundance, ease in growing, documented introduction history and recent evolutionary response to new environments (e.g. herbicide resistance). Since that time, considerable effort has been applied to identify the

Correspondence: Kathryn Hodgins, Fax: +61 3 9905 5613; E-mail: kathryn.hodgins@monash.edu

ecological factors contributing to invasion success (e.g. Catford *et al.* 2009; Lockwood *et al.* 2009), but more recently there has been renewed interest in understanding the evolutionary changes associated with invasion (Dlugosch & Parker 2008; Hodgins & Rieseberg 2011; Colautti & Barrett 2013; Turner *et al.* 2014). Now that genomic information is becoming abundant for a wide array of nonmodel organisms, including invasive species, we can begin to uncover the genomic causes and consequences of biological invasions.

A comparative genomics approach to invasive species research has the potential to identify adaptive genetic changes that are commonly associated with the evolution of invasive species, as well as those that are idiosyncratic. This tactic may also help pinpoint which changes were selected for after introduction, and which have evolved in the native range to predispose certain groups to become problematic invaders. Such information could help elucidate why invasive species are abundant in some lineages but not in others (Kuester *et al.* 2014). The identification of selected changes in invasive populations with independent origins will provide evidence for common genetic responses to colonization. Similarly, finding adaptive responses in the same functional groups of genes among invaders might reveal shared trade-offs that contribute to invasion success, even if the particular genes involved differ among populations or species (Lai *et al.* 2008; Mayrose *et al.* 2011; Guggisberg *et al.* 2013; Hodgins *et al.* 2013b).

While positive selection may play an important role in the evolution of invasive species, introduced populations often experience founder events and an increased likelihood of repeated bottlenecks (Barrett & Shore 1989; Dlugosch & Parker 2008). Because of these processes, significant losses of both allelic richness and heterozygosity in introduced populations are common and gains of diversity are infrequent (Dlugosch & Parker 2008). Population expansion during the spread of an invader across a landscape can have important consequences for genetic variation and fitness. As a species increases its range, genetic drift is predicted to be stronger at the leading edge of the expanding wave front, due to low population densities (Slatkin & Excoffier 2012; Peischl *et al.* 2013). This can have consequences for spatial patterns of neutrally evolving variants (Klopfstein *et al.* 2006; Slatkin & Excoffier 2012), as well as those under selection (Lehe *et al.* 2012; Peischl *et al.* 2013). Genetic drift on the leading edge of range expansions can result in a long-lasting accumulation of deleterious mutations over most of a species range. A recent study revealed that population expansion has left its imprint on the spatial distribution of neutral and deleterious alleles in humans (Peischl *et al.* 2013). Similarly, organisms that have experienced considerable expansion in range size due to their association with humans, such as weeds, may be expected to bear a significant expansion load, yet this remains to be tested.

The Asteraceae family, home to many of the world's worst weeds and invasive species (Baker 1974), is an ideal system for a comparative genomics approach to invasive species research. For example, 588 species in the Asteraceae are considered globally invasive including two of the 33 plants listed by the IUCN as the 100 World's Worst Invasive Alien Species, and 94 species on the US Federal and state noxious weed list (Chamberlain & Szöcs 2013; EOL 2014; GISD 2014; USDA, NRCS 2014). Furthermore, extensive genomic resources have been developed for this family through the Compositae Genome Project (www.cgpdb.ucdavis.edu). These include transcriptome assemblies from upwards of 40 different species, many of which are crops, feral weeds, crop wild relatives and invasive species (Barker *et al.* 2008; Lai *et al.* 2012; Scaglione *et al.* 2012; Hodgins *et al.* 2013b), as well as assemblies of introduced and native genotypes from six invasive species and four out groups to the family. Although the traits associated with successful Asteraceae invasive species vary across taxa (Muth & Pigliucci 2006), herbicide resistance (Holt *et al.* 2013), as well as growth–defence/stress tolerance trade-offs are commonly observed in weedy species in this family (Hodgins & Rieseberg 2011; Mayrose *et al.* 2011; Guggisberg *et al.* 2013; Turner *et al.* 2014) including those that are the subject of this study.

Here we implement a comparative genomics approach to examine changes in protein evolution associated with invasion in Asteraceae species at both micro- and macroevolutionary scales using transcriptome data. First, we asked whether there was evidence of recent parallel evolutionary change among species during invasion by comparing native and introduced genotypes from six focal species. We tested for positive selection in introduced genotypes by identifying genes with elevated nonsynonymous nucleotide substitutions (dN) relative to synonymous substitutions (dS). Following this, we took a broader approach and identified rapidly evolving genes across the family and specifically in weedy lineages. This allowed us to ascertain the genes that may contribute to the propensity of certain groups to become invasive. We then examined changes in genome-wide rates of deleterious mutation to assess whether there was a shift in the genetic load associated with introduction.

Materials and methods

Data set

We used previously published de novo transcriptome assemblies (Barker *et al.* 2008; Lai *et al.* 2012; Scaglione

et al. 2012; Hodgins *et al.* 2013b) for 39 species, which include 35 Asteraceae and four out group taxa (Table S1, Supporting information; www.cgpdb.ucdavis.edu). We classified each of the species as invasive or noninvasive (see Table S1, Supporting information) using the Encyclopedia of Life invasive species comprehensive list, which was accessed programmatically on August 12, 2014 using the taxize package in R (Chamberlain & Szöcs 2013). For some of the species represented in our data set, multiple transcriptomes were available. We included these samples in our ortholog identification and preferentially selected 454 and Illumina transcriptomes as they were generally more complete than Sanger assemblies. For *Helianthus annuus* (annual sunflower), *Ambrosia trifida* (giant ragweed), *A. artemisiifolia* (common ragweed), *Centaurea diffusa* (diffuse knapweed), *C. solstitialis* (yellow starthistle) and *Cirsium arvense* (Canada thistle), de novo assemblies from native and introduced samples were available. From here on, we refer to these six species as our 'focal invasive taxa' (see Appendix S1, Supporting information for species descriptions). The six focal species were used for micro-evolutionary comparisons and the 35 species as well as the out groups were used for the macro-evolutionary comparisons.

Ortholog identification and alignments

For each transcriptome, we removed redundant transcripts, representing alternatively spliced transcripts, alleles or close paralogs, by clustering using Cd-Hit-Est (94% identity, word size = 8 and both strands were compared; Li & Godzik 2006; Fu *et al.* 2012). A single representative sequence from each cluster was retained and used for further analyses. We identified the most likely open reading frames for all orthologs using Transdecoder (option –search_pfam; Haas *et al.* 2013). Open reading frames were translated and annotated through BLASTP to the TAIR 10 database (The Arabidopsis Information Resource; arabidopsis.org). To validate the predicted open reading frames from Transdecoder, we only retained those with a pfam hit or hit to *A. thaliana*. On average, this resulted in 20% of the predicted proteins being removed from the analysis. Confirmation of the predicted proteins reduced the possibility that complete and partial open reading frames were misidentified, reducing a potential source of error, although we acknowledge that we are likely missing important loci that have no homology to proteins in these databases. We conducted an all-against-all BLASTP (e-10). Using these results, we identified orthologs with ORTHAGOGUE version 1.0.2 applying the bit score option and 50% overlap (Li *et al.* 2003). We performed the ortholog identification for all transcriptomes across the entire tree, as well as for subsets of the data used in further analyses. These subsets consisted of (i) species in Asteroideae and Carduoideae, the two subfamilies that contain our focal invasive taxa, (ii) the six focal invasive taxa, and (iii) triplets representing one transcriptome for each introduced and native genotype of our focal invasive taxa, and a third transcriptome from a closely related noninvasive out group (cf. triplet analyses below). For all comparisons, we only used one-to-one orthologs.

We extracted the predicted coding sequences for each transcript and aligned the nucleotide sequences for each orthogroup using Prank (+F option). We used the codon model for the alignments, which is preferred over the amino acid model (Löytynoja & Goldman 2008). We chose Prank because it takes evolutionary relationships into account, outperforming other alignment programs (Löytynoja & Goldman 2008; Fletcher & Yang 2010; Markova-Raina & Petrov 2011; Jordan & Goldman 2012). To automatically remove any sequences resulting in alignment errors, we used Guidance (Penn *et al.* 2010), which generates replicate alignments using a slightly perturbed guide tree with Prank (amino acid model, 30 bootstraps) as the bootstrap aligner, for those alignments with at least four or more sequences. We used a conservative Guidance sequence score cut-off of 0.9 and repeated the Prank alignments. Additionally, to remove potential paralogs, we used a tree-based approach (see Appendix S1, Supporting information).

Species tree construction

We identified the 70 orthogroups with the fewest missing taxa and concatenated the sequences for each species. We visually inspected these data as well as the individual gene trees and removed any sequences representing likely paralogs. We constructed a maximum-likelihood tree in RAXML version 8.0.6 (Stamatakis 2006) using the GTR+G model, which was selected based on JMODELTEST 2.1.4 results (Darriba *et al.* 2012). We did not include gene partitions to prevent overparameterization of the model. The tree was rooted using four out-groups to the Asteraceae family. We used the resulting tree topology for the downstream analyses of positive selection.

Pairwise comparisons of native and introduced transcriptomes

For all orthogroups for which a native and introduced genotype of our focal invasive taxa had one-to-one orthologs, we conducted pairwise comparisons to determine divergence at nonsynonymous and synonymous sites in the coding sequence using PAML version 4.5 (Yang 1997, 2007; see Appendix S1, Supporting

information). Using orthogroups represented across multiple native and introduced pairs, we examined whether patterns of divergence were conserved within and among species by determining whether there was a significant positive correlation in dN/dS ratio (Spearman's rho). Positive correlations would indicate consistency in the strength of selection (purifying and/or positive) among species.

Site-specific positive selection

We conducted an analysis of site-specific tests of positive selection using PAML 4.5 to identify genes that were rapidly evolving across the family, and within the two subfamilies (Asteroideae and Carduoideae) that contain our focal weedy species. For the family-wide analysis, we examined orthogroups with at least three species present in the Asteroideae, Carduoideae and Cichorioideae subfamilies. For the Asteroideae and Carduoideae analysis, only orthogroups with at least four species in each group were used. We trimmed and unrooted the maximum-likelihood species tree to the species in each alignment, using the R package Ape (Paradis *et al.* 2004). We then applied the site model in CODEML to estimate dN and dS at each codon averaged across all branches in the tree. We tested for sites evolving by positive selection by comparing M1a (nearly neutral) and M2a (positive selection), and M7 (beta) against M8 (beta & ω) (F3X4, and transition/transversion ratios estimated). Twice the difference in log-likelihood values M7:M8 (2 d.f.) and M1a:M2a (2 d.f.), comparisons were assessed for statistical significance using the χ^2 distribution. Because the M8 model can be influenced by starting parameters, we ran the program using different starting values of ω (0.4 and 1.5). To statistically minimize the false discovery rate (FDR), we compared *P*-values to critical values calculated based on $\alpha = 0.05$ (p.adjust package in R; Benjamini & Hochberg 1995).

PAML branch and branch-site models

While the site-specific analyses identify genes that are rapidly evolving at specific amino acids across the tree, they do not identify changes in the evolutionary rate that are specific to particular branches. To do this, we implemented PAML's branch and branch-site models for two data subsets. The first consisted of triplets of sequences represented by native and introduced transcriptomes for each of our six focal invasive species and a closely related noninvasive out group. The out groups we used were *Parthenium argentatum* for ragweeds, *Carthamus palaestinus* for *C. diffusa*, *C. solstitialis* and *C. solstitialis*, and *Echinacea angustifolia* for *H. annuus*. In each of these triplets, we marked the introduced genotype as foreground branch (i.e. the branch that is tested for evidence of positive selection). The second data subset consisted of native and introduced genotypes for the six focal invasive taxa. For this analysis, we retained orthogroups containing at least two native-introduced pairs and marked all introduced genotypes as foreground branches.

Branch models allow ω to vary among branches but not among sites and therefore may detect positive selection in specific genes along foreground lineages. We compared the null model, which estimates one ω for all branches, with the alternative model, which estimates one ω for the foreground branches and one ω for the background branches. To identify codons that display evidence of positive selection in specific genes and along foreground branches, we used PAML's branch-site models. We compared the null model, which fixes $\omega_2 = 1$ (Zhang *et al.* 2005) with the alternative model, in which ω_2 is estimated ($\omega_2 \geq 1$). To avoid the detection of local peaks, we ran the branch-site models using different starting ω values (ω = 0.5, 1, 1.5, 2). Other parameters as well as the significance were assessed similarly to the site models, but using one d.f. for likelihood ratio tests (see Appendix S1, Supporting information).

Stochastic branch-site models

We also implemented branch-site models for positive selection using Fitmodel (Guindon *et al.* 2004). The Fitmodel analysis was used for the family-wide data set, as well as for the Asteroideae and Carduoideae subfamilies. We limited our analysis to alignments with six or more sequences due to computational time and because the power to detect selection diminishes when there are few species in the tree (Lu & Guindon 2014). Switching models allow each codon site to change the selective regime and thus be affected by selective pressures at different time points. To test for evidence of positive selection varying down branches, we compared the M1a model (no positive selection) to the M2a model, which included selection. In both cases, switching was allowed to occur. Sites with episodes of positive selection ($P > 0.90$) were detected a posteriori using a Bayesian approach (Guindon *et al.* 2004). The likelihood ratio test statistic with one d.f. was assumed based on results of simulations (Lu & Guindon 2014). Similarly to the PAML analysis, all putatively selected sites were confirmed by visually inspecting the alignments.

Functional annotation and GO analysis

Using the results of the BLASTP to the TAIR 10 database (cf. above), we selected the top hit for each species assigned to each orthogroup. We were specifically interested in examining positive selection in target-site

resistance (TSR) genes for several herbicides (Table S2, Supporting information), due to the repeated evolution of herbicide resistance in Asteraceae weeds (Holt *et al.* 2013). We conducted a BLASTP against the predicted proteins for each species and then identified putative TSR orthogroups using these annotations. We assigned GO terms to each orthogroup based on the GO *A. thaliana* mappings to the top hits and performed a GO enrichment analysis using topGO and the parent–child method (Alexa *et al.* 2006; Grossmann *et al.* 2007; see Appendix S1, Supporting information). To identify broader patterns, we then identified GO slim terms for each gene using the TAIR 10 database and conducted a chi-squared test to compare the number of loci found in each category for the background and selected loci.

Identification of deleterious alleles

Using the pairwise alignments, we wrote custom scripts to identify if SNPs caused amino acid changes between the introduced and native transcriptomes for each of the six focal invasive species. The deleterious effects of amino acid substitutions were then predicted for proteins derived from each gene with Provean (Choi *et al.* 2012). Provean uses homologous sequences identified by PSI-BLAST against protein databases (nr protein database) and identifies changes in sequence similarity of a query sequence to the protein sequence homolog by comparing an alignment-based score before and after the amino acid change to the query sequence. We used the recommended threshold of −2.5 for the identification of deleterious sites using the nr database and a paired *t*-test to determine whether there were differences in the number of deleterious amino acids separating each native and introduced comparison. We then identified out groups using alignments within each subfamily to determine derived amino acid changes and eliminated any sites where both amino acids were found in the out group species as these were potential paralogs. We applied a Fisher's exact test to examine differences in the proportion of derived deleterious amino acids between the native and introduced transcriptomes for each of the six species and then assessed overall significance using a paired *t*-test. If multiple comparisons could be conducted within a species, we took the average for each range.

Results

Pairwise comparisons of native and introduced transcriptomes

Our pairwise comparison of dN/dS ratios between native and introduced transcriptomes, performed using 2836–8368 total orthogroups depending on the focal species considered, identified between 0.6% and 2.6% of loci with elevated evolutionary rates (Table S3, Supporting information). All rapidly evolving genes were unique to each species, except one that was shared between *C. arvense* and *A. trifida*, for which no homolog in *A. thaliana* was identified. However, using the all-against-all blasts, we did identify 20 overlapping homolog groups (i.e. where the reciprocal blast hits were identified) among species for rapidly evolving genes (Table S3, Supporting information). For the conserved orthologs identified among the focal species, we found significant correlations between species, except in some comparisons with *A. trifida*, as well as stronger correlations within species, where multiple comparisons could be conducted (Fig. 1). When multiple native and introduced comparisons were possible within each species, we tested all combinations and present average values in Fig. 1.

Among the quickly evolving genes identified, there were 21 GO terms overrepresented for *A. artemisiifolia*, including defence response (four genes), 27 in *A. trifida*, nine in *C. diffusa*, 24 in *C. solstitialis*, seven in *C. arvense* and three terms in *H. annuus* (Table S3, Supporting information; $P < 0.05$). However, the GO analyses of the specific terms did not reveal any significant overrepresentation of terms after correcting for multiple tests for

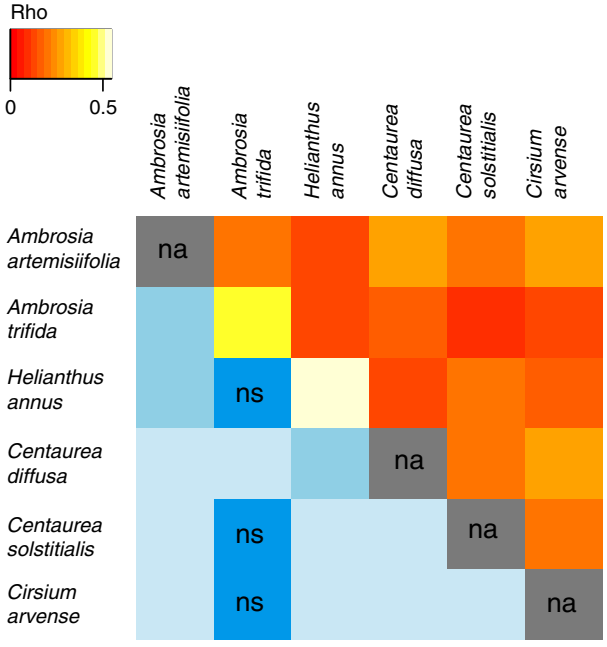

Fig. 1 The Spearman correlation coefficient for dN/dS of pairwise comparisons between native and introduced samples for six species (upper triangle) and their significance (lower triangle). Multiple comparisons within a species were available in some cases (diagonal). $P > 0.05$ (dark blue), $P < 0.05$ (medium blue) and $P < 0.001$ (light blue).

Table 1 Results of chi-squared tests comparing the number of genes in each GO slim category for genes showing evidence of positive selection in pairwise comparisons between native and introduced genotypes relative to those that did not

Species	χ^2	d.f.	P	Top three GO slim categories overrepresented
Ambrosia artemisiifolia	52.59	43	0.15	
Ambrosia trifida	58.23	43	0.06	
Centaurea diffusa	41.95	43	0.52	
Centaurea solstitialis	64.29	43	0.02	Other binding, other enzyme activity, other cellular components, cell wall
Cirsium arvense	70.58	43	0.005	Unknown cellular components, other metabolic processes, transcription factor activity
Helianthus annuus	67.54	43	0.009	Hydrolase activity, response to stress, cell wall

any of the pairwise comparisons. Examination of the GO slim terms identified significant differences in the distribution of terms in *C. solstitialis*, *C. arvense* and *H. annuus* (Table 1).

PAML *branch and branch-site models*

The final tree topology (Fig. 2) of our species tree generally agreed with the previously published Asteraceae supertree (Funk *et al.* 2005). The branch and branch-site models for the triplet data sets, performed using 939–2590 total orthogroups depending on the focal species considered, revealed evidence of positive selection in <1.28% and 5.75% of orthogroups, respectively, for each focal species (Table S4, Supporting information). For those genes where orthology could be identified, none of the rapidly evolving genes were shared across species. Using the all-against-all blasts, we did identify 24 overlapping homolog groups among two to three species (Table S4, Supporting information). The corresponding TAIR hits and topGO results for these orthogroups are given in Table S4 (Supporting information). The branch model performed for all six focal species, which considered information from 1792 total orthogroups, concomitantly revealed evidence of positive selection in three orthogroups (Table 2 and Table S5, Supporting information). The branch-site model implemented for the same data set, however, did not identify any significant orthogroups.

Site-specific positive selection

For the M1a:M2a comparison, after FDR correction, we found evidence of positive selection in one of the 397 (0.003%; Table S6, Supporting information) orthogroups across the entire family. Similarly, for Carduoideae and Asteroideae, 32 of the 7654 (0.42%) and 16 of the 12,416 (0.13%) orthogroups showed evidence of positive selection, respectively. For the M7:M8 comparison, we found evidence for positive selection in six genes (0.02%) across the entire family, in 53 genes (0.69%) for Carduoideae and in 55 genes (0.44%) for Asteroideae. Most of these genes were also identified using the M1a:M2a models.

For the Carduoideae, we found 10 GO terms overrepresented at $P < 0.05$ including response to biotic stimulus, vegetative phase change, response to UV, response to wounding and response to virus (Table S6, Supporting information). For the Asteroideae, we identified 40 GO terms overrepresented at $P < 0.05$ including GO terms related to response to abiotic stimulus and cell wall biogenesis (Table S6, Supporting information). However, these terms were not significant after FDR correction. When comparing the GO slim categories between background and selected genes, the Carduoideae was marginally significantly different ($\chi^2 = 58.63$, d.f. = 43, $P = 0.056$; with *other molecular functions*, *binding*, *enzyme activity* and *response to abiotic or biotic stimulus* the most overrepresented terms) and the Asteroideae was not significantly different ($\chi^2 = 46.35$, d.f. = 43, $P = 0.34$).

Stochastic branch-site models

Across the entire family, we found evidence for positive selection in 29 of the 397 genes (M1a+S1:M2a+S1; Table S6, Supporting information) after FDR correction (7.8%). All genes possessed at least one site showing evidence of positive selection (Bayesian posterior probability >0.90) in branches leading to at least one introduced species (Table S6, Supporting information; for designations, see Table S1, Supporting information). Within the Carduoideae, we found evidence of positive selection in 55 of the 2259 genes after FDR correction (2.4%). Of those genes, 46 had at least one site showing evidence of positive selection (Bayesian posterior probability >0.90) and 44 of those showed evidence of positive selection in branches leading to at least one introduced species (*Centaurea*, *Cynara* and *Cirsium*) with two additional loci showing evidence of positive selection in *Carthamus* alone. We identified significant evidence for positive selection in 212 of the 6356 genes in the Asteroideae (3.3%). Of those genes, 83 had at least one site showing evidence of positive selection (Bayesian posterior probability >0.90) and all but three of those showed evidence of positive selection in branches leading to at

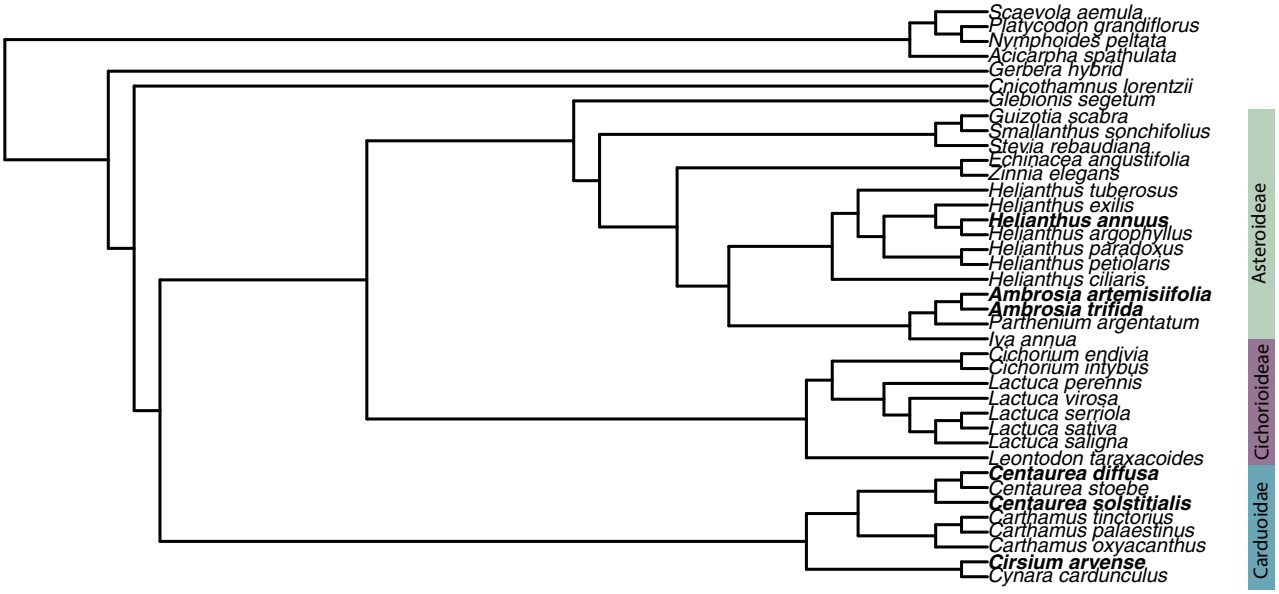

Fig. 2 The tree topology used in the PAML and Fitmodel analysis. Bold font is used to indicate our focal invasive taxa.

Table 2 PAML results for three orthogroups that showed evidence of significant changes in dN/dS ratio (ω) in native vs. introduced genotypes using branch models

Orthogroup	TAIR ID	Description	Significant foreground branches	Likelihood ratio†
Ortho group4063	AT5G42990.1	Ubiquitin-conjugating enzyme 18	*A. artemisiifolia, H. annuus*	27.58***
Ortho group5021	AT5G34930.1	Arogenate dehydrogenase	*A. trifida, H. annuus*	17.05*
Ortho group5575	AT3G12120.1	Fatty acid desaturase 2 (FAD2)	*A. trifida, H. annuus*	53.06***

†Significance after FDR correction. *$P < 0.05$;***$P < 0.001$.

least one introduced species (*A. artemisiifolia*, *A. trifida*, *G. segetum*, *H. annuus*, *H. ciliaris*, *H. petiolaris*, *H. tuberosus* or *I. annua*).

We then identified the putative function of the significant genes using the BLASTP results to the *A. thaliana* proteins (Table S6, Supporting information). Across the entire family, we found five GO terms overrepresented ($P < 0.05$; Table S6, Supporting information). For the Carduoideae, we found 14 GO terms overrepresented at $P < 0.05$ including response to herbicide (one gene; Table S6, Supporting information). For the Asteroideae, we identified 20 GO terms overrepresented at $P < 0.05$ including GO terms related to response to osmotic stress (seven genes), response to gibberellin-mediated signalling pathway (two genes) and reproductive process (12 genes; Table S6, Supporting information). However, these GO terms were not significant after FDR correction. The analysis of the GO slim terms revealed a marginally significant difference in the proportion of genes in the different categories between the background and selected loci across the family ($\chi^2 = 54.74$, d.f. = 25, $P = 0.10$) with *other enzyme*, *response to abiotic or biotic stimulus* and *response to stress* as the most overrepresented GO slim terms. There was no significant difference in the Asteroideae ($\chi^2 = 43.18$, d.f. = 43, $P = 0.46$) or Carduoideae ($\chi^2 = 30.5$, d.f. = 43, $P = 0.92$).

Herbicide resistance target genes

We identified filtered orthogroup alignments homologous to several TSR genes (Table S2, Supporting information). Homologs of α-tubulin, β-tubulin, and ALS were found in the Carduoideae and Asteroideae alignments as well as psbA and PPX2L in the Asteroideae alignments. For both the PAML site model M1a:M2a and Fitmodel analysis, we identified ALS as significant in the Asteroideae ($q < 0.05$ after correcting for multiple comparison for the TSR genes only). Comparisons of the sites putatively under selection (amino acids 484 and 598 in the *A. thaliana* protein) were not known amino acids conferring resistance in other species (see

www.weedscience.org/Mutations/MutationDisplayAll.aspx), and those known to confer resistance were not represented in our alignments for that gene. Several homologs of α-tubulin and β-tubulin were identified, and both models of site-specific positive selection in PAML were significant for a predicted β-tubulin in the Carduoideae ($q < 0.05$ as significant after correcting for multiple comparison for the TSR genes).

Proportion of deleterious alleles

We found no effect of introduction on the number of deleterious alleles among the six focal species ($t_5 = 0.52$, $P = 0.63$). The same was true when we restricted our analysis to the proportion of deleterious derived alleles in each of the transcriptomes. Individual Fisher's exact tests of the paired samples demonstrated differences in the proportion of derived deleterious alleles for different species: a significantly greater proportion of deleterious variants in the introduced sample from Australia in *H. annuus*, no significant difference in *A. artemisiifolia* and some *A. trifida* comparisons, and significantly lower number of deleterious variants in the introduced range for the remaining comparisons (Fig. 3; Table S7, Supporting information). A paired *t*-test of the per cent difference between the native and introduced transcriptomes revealed no general pattern ($t_5 = -0.82$, $P = 0.45$).

Discussion

Positive selection in introductions

To test whether there were common adaptive changes at the genetic level, we examined evidence for positive selection in introduced genotypes across multiple species. Detecting the genomic signature of selection has an advantage in that no a priori determination of the traits under selection is required. However, common garden studies comparing native and introduced populations of *A. artemisiifolia*, *C. solstitialis*, *C. arvense* and *C. diffusa*, as well as comparisons between weedy and wild *H. annuus* have been conducted and striking similarities have emerged (Eriksen *et al.* 2012; Guggisberg *et al.* 2013; Hodgins *et al.* 2013a; Turner *et al.* 2014). In these studies, introduced populations tended to be larger with higher reproductive output and delayed reproduction relative to native populations. Introduced populations of all three species also show evidence of reduced abiotic tolerance and in particular are more susceptible to drought; comparisons of weedy and wild *H. annuus* show a similar pattern (Mayrose *et al.* 2011). In all four species, there is limited evidence for tradeoffs of competitive ability with defence, although in most cases the effects of specialist herbivores have not been examined (but see Turner *et al.* 2014). These parallel evolutionary responses to introduction suggest that there may be replicated genetic changes during introduction among these species, either in terms of particular genes under selection or the functional groups commonly targeted.

We found little evidence of repeated adaptive change among any of the six focal species investigated in this study. While, in most cases, dN/dS ratios were correlated across species, this appeared to be due to consistent differences in the strength of purifying selection among loci (Fig. 1). Our analyses identified many genes showing evidence of positive selection. However, there was little overlap among species in the identity of positively selected orthologs. For example, only three genes,

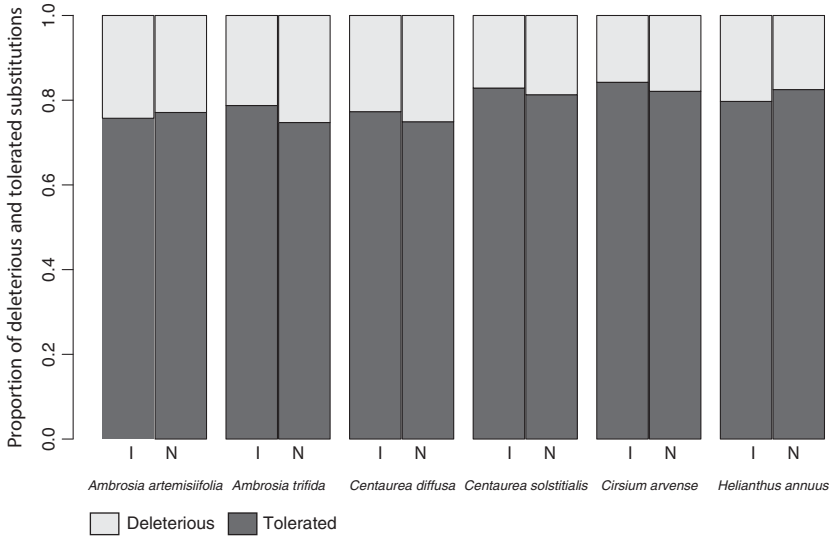

Fig. 3 The proportion of derived deleterious alleles in native (N) and introduced (I) samples across six species identified using Provean. Average values are presented for species with multiple native and introduced samples.

including one (FAD2) implicated in seed dormancy, as well as seed oil content during sunflower domestication (Linder 2000; Chapman & Burke 2012), have evidence for repeated adaptive evolution using our branch models, but these patterns are only replicated in two species (Table 2). Comparisons of genes inferred to be under positive selection during domestication in this family have found little commonality between any two domestication events using this approach (Kane *et al.* 2011). Notably, this pattern was found despite similar selective pressures and strong artificial selection across thousands of years. Given the shorter timescale and the more diverse ecological conditions likely encountered during invasions relative to an agricultural environment, the fact that there is little evidence of repeatability in positively selected genes in our comparisons between native and introduced genotypes is not surprising. Whether the lack of repeatability at the genetic level during invasion will be found across other taxonomic groups remains to be seen. However, given the variable evidence for consistent phenotypic change during invasion (e.g. Bossdorf *et al.* 2005; Colautti *et al.* 2009), and the complex genetic architecture of many of the putative traits under selection (e.g. plant growth and defence response), we predict that our findings are a harbinger of what is to come when comparing multiple species.

We identified evidence that genes in the same family were evolving rapidly across species, suggesting that at least in some cases, similar types of genes were under selection. These genes have diverse functions in *A. thaliana*, including defence response (e.g. AT4G16890, AT3G15850, AT2G25620), abiotic stress response particularly osmotic stress (e.g. AT5G59320) and flavonol biosynthesis (AT5G54160). Although there was limited evidence for the overrepresentation of functional categories in the divergent genes for each species, we found many putatively selected loci important for abiotic and biotic responses. For example, in pairwise comparisons of *A. artemisiifolia*, four genes were annotated as defence response genes, and in *H. annuus*, five genes were annotated as response to stress and a further four were annotated as response to abiotic or biotic stimulus. These GO slim categories were some of the most overrepresented in pairwise comparisons of *C. arvense*. In microarray experiments of *C. arvense* and *A. artemisiifolia* (Guggisberg *et al.* 2013; Hodgins *et al.* 2013a), differences in gene expression between the native and introduced range have also involved stress response genes and, in particular, genes associated with secondary metabolism and detoxification such as the cytochrome P450 gene family. Although differences in expression are not necessarily expected to coincide with changes in coding sequence (e.g. Renaut *et al.* 2012; but see Chapman *et al.* 2013), we identified one differentially expressed gene and putatively selected loci in *A. artemisiifolia*, a UDP-glucosyl transferase (AT1G22340). These genes are involved in metabolism of plant hormones, all major classes of plant secondary metabolites and xenobiotics such as herbicides (Ross *et al.* 2001). Analyses of gene expression between native and introduced populations are underway in the remaining species, and comparisons will offer additional insights into the repeatability of molecular changes during invasion.

The power of our approach relies on our ability to examine large numbers of genes across multiple species to determine commonalities in the signature of selection. However, we lack sufficient sampling within each species to determine whether the observed differences are consistently associated with introductions within each species. Large-scale population genomic sampling for species is underway using a variety of techniques (e.g. transcriptome resequencing, GBS and RAD tags), and these data will be important to determine whether the genes and functional groups that we have identified have likely diverged as a result of selective processes during introduction and spread. Some of the candidate loci could be due to mistakenly identifying close paralogs as orthologs, although all of the orthologs that we identified showed high sequence similarity (Table S3, Supporting information). Other processes may be maintaining variation within and among populations of these species that are not associated with invasion. For example, we might expect that some of the genes identified as highly divergent, such as disease resistance genes, are maintained by balancing selection within the species generally (Bergelson *et al.* 2001). The recent occurrence of the introductions (100–200 years) and the rapid evolutionary change needed for this approach to be successful, where multiple amino acid changes are often required to identify diverging genes, suggests that standing variation segregating in the initial introductions would likely be the origin of the putatively selected loci (Prentis *et al.* 2008). In contrast, any divergence between the native and introduced ranges that is due to de novo mutations in the introduced range is unlikely to be identified using this approach, given the time needed for new mutations to arise.

Rates of positive selection in the Asteraceae

We found that a small fraction of genes exhibited significant evidence of positive selection using the site models (<1%), while the stochastic branch-site models identified many more candidate loci (2–8%). Site models are mainly able to detect recurrent diversifying selection and lack the capacity to detect episodic adaptive

evolution. Our estimates of site-specific positive selection are lower than previous findings (~4% for annual *Helianthus* and *Lactuca* species vs. <1% for each subfamily; Kane *et al.* 2011). This could be in part because of the broader phylogenetic scale that we were examining and suggests that a particular site is rarely the subject of selection over longer evolutionary timescales. In keeping with this, our stochastic branch-site models identified a higher proportion of candidate loci and the proportion was greater in the analysis across the family relative to the subfamilies, as more diverse taxa were added. In some cases, the Bayesian posterior analysis of the stochastic branch-site models did find evidence that the same gene was the target of selection across the entire tree, but the particular sites would often vary among lineages. In other cases, selection was clearly episodic or recurrent down a particular section of the tree.

However, both approaches likely considerably underestimate the actual rate of positive selection for a number of reasons. First, by identifying orthology using a blast-based approach, we are potentially missing rapidly evolving genes that are highly divergent. Moreover, missing or incomplete transcripts are common in some of the assemblies and limited our ability to assess selection across the entire set of species or proteins in many cases. Second, by necessity, we restricted our analysis to one-to-one orthologs, which likely biases our sample towards single-copy genes that can be subject to stronger purifying selection than multicopy gene families (Waterhouse *et al.* 2011). Ortholog identification may have particularly impacted our ability to detect selection at broader phylogenetic scales. Third, we rigorously trimmed and filtered regions that were difficult to align and may harbour diverging sections of the gene. However, conserved regions might be more likely to contain positively selected substitutions (Bazykin & Kondrashov 2012) and removing problematic regions that generate false positives should improve our capacity to detect positive selection (Jordan & Goldman 2012; Privman *et al.* 2012).

Genes under selection in the Asteraceae

Many of the genes that appear to be rapidly evolving in the Asteraceae appear to be likely targets of positive selection in plants in general. Although our gene ontology analysis did not reveal a strong pattern in the types of genes overrepresented in the set of positively selected loci (Table S6, Supporting information), in most cases, the number of selected loci was small which limited the power of the test. In plants, defence genes, in particular leucine-rich repeats of R genes, provide some of the best examples of co-evolutionary 'arms races' resulting in rapid evolutionary change (Bergelson *et al.* 2001). We identified many defence-related genes, particularly in the Carduoideae, including ENHANCED DISEASE SUSCEPTIBILITY 1 (AT3G48090), which impacts salicylic acid (SA) levels and enhances susceptibility to pathogen infection (Falk *et al.* 1999; Nawrath *et al.* 2002), CUCUMOVIRUS MULTIPLICATION 1 (AT4G18040) and BINDING TO TOMV RNA 1 (AT4G37760), both of which are involved in viral resistance (Gao *et al.* 2004; Fujisaki & Ishikawa 2008; Contreras-Paredes *et al.* 2013). Many other stress response genes were identified, including GRX480, a member of the glutaredoxin family that is induced by SA (Ndamukong *et al.* 2007), heat-shock proteins (e.g. AT5G02500 AT3G47650), as well as glutathione S-transferases (e.g. AT2G30860, AT1G78380), involved in abiotic and biotic stress response, including drought stress and herbicide resistance (Rouhier *et al.* 2008; Powles & Yu 2010). In the Asteroideae, abiotic stress response genes predominated, including seven genes related to osmotic stress response, such as LOW EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 2 (AT2G36530; Ishitani *et al.* 1997) and a glycine-rich RNA-binding protein (AT3G23830), important for cold tolerance and osmotic stress response in *A. thaliana* (Kwak *et al.* 2011). Taken together, these data indicate that genes important for environmental responses appear to be evolving rapidly in this family, which is consistent with previous findings (Kane *et al.* 2011).

Conflicts between genomes can be an important driver of evolutionary rate in plants (e.g. Fujii *et al.* 2011) and this appears to also be the case in our study. We identified candidate genes involved in interactions between the nucleus and either the mitochondrion or the chloroplast including maternal effect embryo arrest proteins (AT3G10110 and AT5G05950; Pagnussat *et al.* 2005). High rates of reproductive protein evolution are also common in both animals and plants (reviewed in Clark *et al.* 2006) and we identified several reproductive genes, particularly in the Asteroideae where 12 were identified. For example, we found several genes involved in male and female gametophyte development (e.g. APK3 AT3G03900, Mob1 AT5G45550 and AT4G30930; Portereiko *et al.* 2006; Mugford *et al.* 2010; Galla *et al.* 2011), which were also localized to the mitochondrial or chloroplast, as well as two putative genes involved in the regulation of flowering time (FCA and CONSTANS-like; Macknight *et al.* 1997; Reeves & Coupland 2001; Hassidim *et al.* 2009).

The evolution of herbicide resistance is increasingly common in weeds (Holt *et al.* 2013) and is known to occur in several species used in this study (e.g. ALS inhibitors in *A. trifida*, *A. artemisiifolia* and *H. annuus*, and resistance to synthetic auxins in *C. arvense* and

C. solstitialis http://www.weedscience.org/Summary/home.aspx), although the resistance status to various herbicides of the genotypes used in this analysis was not known. There were several genes showing evidence of positive selection that play important roles in detoxification and response to xenobiotics (e.g. glutathione S-transferases, Cytochrome p450s; Powles & Yu 2010; Délye *et al.* 2013) and could be important for herbicide response in this family. However, there are many genes that are known to be the direct targets of commonly applied herbicides (i.e. TSR genes). We tested these for evidence of positive selection and found elevated evolutionary rates in two cases (β-tubulin in Carduoideae and ALS in the Asteroideae). Several ALS-resistant weeds species, such as *H. annuus* and ragweeds, are present in our analysis; however, none of the substitutions in these alignments coincided with known resistance alleles, which suggests either the evolution of novel resistance mutations or a separate selective mechanism driving evolutionary change in these loci. Several key TSR genes (e.g. EPSPS, the target of glyphosate) were not found in our alignments; however, studies are ongoing in ragweeds and other Asteraceae (e.g. *Conyza* spp.) to uncover the genetic basis of herbicide resistance and relative importance of TSR and non-TSR mechanisms in the group.

Genetic load in the introduced range

We predicted that the demographic changes, such as repeated bottlenecks and founder events associated with establishment and spread during invasion, would result in higher levels of genetic load in introduced genotypes. However, our estimates of genetic load across the six focal species did not reveal a general increase in the proportion of deleterious variants in introduced samples. One of the comparisons (native vs. introduced *H. annuus* from Australia) revealed a significantly greater proportion of deleterious alleles in the introduced genotype, suggesting that this population may have been subjected to higher levels of drift. An isozyme study of genetic diversity in Australian *H. annuus* populations, where it was likely introduced as an ornamental, revealed high levels of variation within, but also among populations, as well as an apparent lack of isolation by distance; this pattern is consistent with multiple distinct source populations (Dry & Burdon 1986). However, the study also found relatively high inbreeding coefficients, and the authors suggested this was due to small population sizes, founder events and potentially some degree of selfing. This may also explain why we found evidence of higher genetic load in this particular comparison.

In contrast to *H. annuus*, the introduced samples from the other five focal species were from a portion of the species' range where they are considered invasive and form large population sizes. These comparisons showed statistically equivalent levels of load or even the reverse pattern with a significantly higher proportion of deleterious alleles in the native samples. With the exception of *A. trifida*, where no genetic data are yet available, previous studies of the introduction history of all of these species show evidence of multiple introductions and high levels of genetic variation within populations (Sun 1997; Marrs *et al.* 2007; Blair & Hufbauer 2010; Chun *et al.* 2010; Gaudeul *et al.* 2011; Guggisberg *et al.* 2012). *Cirsium arvense*, *A. artemisiifolia* and *C. solstitialis* lack substantial isolation by distance in the introduced range, giving further support for repeated long distance dispersal from multiple source populations as the primary driver of genetic variation within the introduced range. Although drift in colonizing species is predicted to increase genetic load, hybridization and population admixture through multiple introductions might limit its severity. Future analyses of the frequency of these deleterious alleles, the degree of admixture across the landscape and the spatial distribution of genetic load within and between the native and introduced range will be essential to more conclusively assess this hypothesis.

Conclusions

Our micro- and macro-evolutionary comparisons revealed that generally there is little overlap in specific genes that appeared to be under positive selection. In our comparisons of native and introduced genotypes across species, there was little commonality in the particular orthologs under selection, despite similarities in phenotypic response during invasion. In some cases, we found similar types of genes under positive selection across two or three introductions, which opens some possibilities of common responses to invasion at the genomic level, but our data suggest that repeated changes in the same gene are unlikely. At the macro-evolutionary scale, we found a similar pattern, where the sites, genes and functional groups showing evidence of positive selection generally differed between subfamilies. These findings point to the idiosyncratic nature of positive selection, where the targets of selection shift across sites, genes and lineages. Moreover, it suggests that selection pressures may be highly variable and outcomes dependent on historical contingency and species-specific genetic constraints, and highlights the possibility of multiple genetic solutions to the challenge of adapting to similar types of environmental change.

Acknowledgements

We thank RI Colautti, EBM Drummond and I Mayrose for advice; Z Lai, LO Oliviera, KM Dlugosch and M Barker for data contributions; and Armando Geraldes as well as three anonymous reviewers for their insightful comments. Funding was provided by a Natural Sciences and Engineering Research Council (NSERC) Discovery grant to LHR, NSERC Vanier CGS and Killam Doctoral Fellowship to DGB, and a Swiss National Science Foundation grant to MAH.

References

Alexa A, Rahnenführer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, **22**, 1600–1607.

Baker H (1974) The evolution of weeds. *Annual Review of Ecology and Systematics*, **5**, 1–24.

Barker MS, Kane NC, Matvienko M *et al.* (2008) Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Molecular Biology and Evolution*, **25**, 2445–2455.

Barrett SCH, Shore JS (1989) Isozyme variation in colonizing plants. In: *Isozymes in Plant Biology* (eds Soltis D, Soltis P), pp. 106–126. Dioscorides Press, Portland, Oregon.

Bazykin GA, Kondrashov AS (2012) Major role of positive selection in the evolution of conservative segments of *Drosophila* proteins. *Proceedings of the Royal Society B*, **279**, 3409–3417.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, **57**, 289–300.

Bergelson J, Kreitman M, Stahl EA, Tian D (2001) Evolutionary dynamics of plant R-genes. *Science*, **292**, 2281–2285.

Blair AC, Hufbauer RA (2010) Hybridization and invasion: one of North America's most devastating invasive plants shows evidence for a history of interspecific hybridization. *Evolutionary Applications*, **3**, 40–51.

Bossdorf O, Auge H, Lafuma L *et al.* (2005) Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia*, **144**, 1–11.

Catford JA, Jansson R, Nilsson C (2009) Reducing redundancy in invasion ecology by integrating hypotheses into a single theoretical framework. *Diversity and Distributions*, **15**, 22–40.

Chamberlain SA, Szöcs E (2013) Taxize: taxonomic search and retrieval in R [v2; ref status: indexed, http://f1000r.es/24v]. *F1000Research*, **191**, 1–28.

Chapman MA, Burke JM (2012) Evidence of selection on fatty acid biosynthetic genes during the evolution of cultivated sunflower. *Theoretical and Applied Genetics*, **125**, 897–907.

Chapman MA, Hiscock SJ, Filatov DA (2013) Genomic divergence during speciation driven by adaptation to altitude. *Molecular Biology and Evolution*, **30**, 2553–2567.

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE*, **7**, e46688.

Chun YJ, Fumanal B, Laitung B, Bretagnolle F (2010) Gene flow and population admixture as the primary post-invasion processes in common ragweed (*Ambrosia artemisiifolia*) populations in France. *The New Phytologist*, **185**, 1100–1107.

Clark NL, Aagaard JE, Swanson WJ (2006) Evolution of reproductive proteins from animals and plants. *Reproduction*, **131**, 11–22.

Colautti RI, Barrett SCH (2013) Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science*, **342**, 364–366.

Colautti RI, Maron JL, Barrett SCH (2009) Common garden comparisons of native and introduced plant populations: latitudinal clines can obscure evolutionary inferences. *Evolutionary Applications*, **2**, 187–199.

Contreras-Paredes CA, Silva-Rosales L, Daròs J-A, Alejandri-Ramírez ND, Dinkova TD (2013) The absence of eukaryotic initiation factor eIF(iso)4E affects the systemic spread of a Tobacco etch virus isolate in *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions: MPMI*, **26**, 461–470.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

Délye C, Jasieniuk M, Le Corre V (2013) Deciphering the evolution of herbicide resistance in weeds. *Trends in Genetics: TIG*, **29**, 649–658.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431–449.

Dry PJ, Burdon JJ (1986) Genetic structure of natural populations of wild sunflowers (*Helianthus annuus* L.) in Australia. *Australian Journal of Biological Sciences*, **39**, 255–270.

Encyclopedia of Life [EOL] (2014) "Invasive Species Comprehensive List". Available from http://eol.org/collections/55367. Accessed 13 Sept 2014.

Eriksen RL, Desronvil T, Hierro JL, Kesseli R (2012) Morphological differentiation in a common garden experiment among native and non-native specimens of the invasive weed yellow starthistle (*Centaurea solstitialis*). *Biological Invasions*, **14**, 1459–1467.

Falk A, Feys BJ, Frost LN *et al.* (1999) EDS1, an essential component of R gene-mediated disease resistance in *Arabidopsis* has homology to eukaryotic lipases. *Proceedings of the National Academy of Sciences, USA*, **96**, 3292–3297.

Fletcher W, Yang Z (2010) The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Molecular Biology and Evolution*, **27**, 2257–2267.

Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**, 3150–3152.

Fujii S, Bond CS, Small ID (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proceedings of the National Academy of Sciences, USA*, **108**, 1723–1728.

Fujisaki K, Ishikawa M (2008) Identification of an *Arabidopsis thaliana* protein that binds to tomato mosaic virus genomic RNA and inhibits its multiplication. *Virology*, **380**, 402–411.

Funk VA, Bayer RJ, Keeley S *et al.* (2005) Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae. *Biologiske Skrifter*, **55**, 343–373.

Galla G, Zenoni S, Marconi G *et al.* (2011) Sporophytic and gametophytic functions of the cell cycle-associated Mob1 gene in *Arabidopsis thaliana* L. *Gene*, **484**, 1–12.

Gao Z, Johansen E, Eyers S *et al.* (2004) The potyvirus recessive resistance gene, *sbm1*, identifies a novel role for translation

initiation factor eIF4E in cell-to-cell trafficking. *The Plant Journal*, **40**, 376–385.

Gaudeul M, Giraud T, Kiss L, Shykoff JA (2011) Nuclear and chloroplast microsatellites show multiple introductions in the worldwide invasion history of common ragweed, *Ambrosia artemisiifolia*. *PLoS ONE*, **6**, e17658.

Global Invasive Species Database [GISD] (2014) 100 of the world's worst invasive alien species. Available from http://www.issg.org/database/species/search.asp?st=100ss&fr=1&str=&lang=EN. Accessed 13 September 2014.

Gray A, Mack R (1986) Do invading species have definable genetic characteristics? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **314**, 655–674.

Grossmann S, Bauer S, Robinson PN, Vingron M (2007) Improved detection of overrepresentation of Gene-Ontology annotations with parent child analysis. *Bioinformatics*, **23**, 3024–3031.

Guggisberg A, Welk E, Sforza R *et al.* (2012) Invasion history of North American Canada thistle, *Cirsium arvense*. *Journal of Biogeography*, **39**, 1919–1931.

Guggisberg A, Lai Z, Huang J, Rieseberg LH (2013) Transcriptome divergence between introduced and native populations of Canada thistle, *Cirsium arvense*. *The New Phytologist*, **199**, 595–608.

Guindon S, Rodrigo AG, Dyer KA, Huelsenbeck JP (2004) Modeling the site-specific variation of selection patterns along lineages. *Proceedings of the National Academy of Sciences, USA*, **101**, 12957–12962.

Haas BJ, Papanicolaou A, Yassour M *et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, **8**, 1494–1512.

Hassidim M, Harir Y, Yakir E, Kron I, Green RM (2009) Over-expression of CONSTANS-LIKE 5 can induce flowering in short-day grown *Arabidopsis*. *Planta*, **230**, 481–491.

Hodgins KA, Rieseberg L (2011) Genetic differentiation in life-history traits of introduced and native common ragweed (*Ambrosia artemisiifolia*) populations. *Journal of Evolutionary Biology*, **24**, 2731–2749.

Hodgins KA, Lai Z, Nurkowski K, Huang J, Rieseberg LH (2013a) The molecular basis of invasiveness: differences in gene expression of native and introduced common ragweed (*Ambrosia artemisiifolia*) in stressful and benign environments. *Molecular Ecology*, **22**, 2496–2510.

Hodgins KA, Lai Z, Oliveira LO *et al.* (2013b) Genomics of Compositae crops: reference transcriptome assemblies and evidence of hybridization with wild relatives. *Molecular Ecology Resources*, **14**, 166–177.

Holt JS, Welles SR, Silvera K *et al.* (2013) Taxonomic and life history bias in herbicide resistant weeds: implications for deployment of resistant crops. *PLoS ONE*, **8**, e71916.

Ishitani M, Xiong L, Stevenson B, Zhu JK (1997) Genetic analysis of osmotic and cold stress signal transduction in *Arabidopsis*: interactions and convergence of abscisic acid-dependent and abscisic acid-independent pathways. *The Plant Cell*, **9**, 1935–1949.

Jordan G, Goldman N (2012) The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Molecular Biology and Evolution*, **29**, 1125–1139.

Kane NC, Barker MS, Zhan SH, Rieseberg LH (2011) Molecular evolution across the Asteraceae: micro- and macroevolutionary processes. *Molecular Biology and Evolution*, **28**, 3225–3235.

Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, **23**, 482–490.

Kuester A, Conner JK, Culley T, Baucom RS (2014) How weeds emerge: a taxonomic and trait-based examination using United States data. *The New Phytologist*, **202**, 1055–1068.

Kwak KJ, Park SJ, Han JH *et al.* (2011) Structural determinants crucial to the RNA chaperone activity of glycine-rich RNA-binding proteins 4 and 7 in *Arabidopsis thaliana* during the cold adaptation process. *Journal of Experimental Botany*, **62**, 4003–4011.

Lai Z, Kane NC, Zou Y, Rieseberg LH (2008) Natural variation in gene expression between wild and weedy populations of *Helianthus annuus*. *Genetics*, **179**, 1881–1890.

Lai Z, Kane NC, Kozik A *et al.* (2012) Genomics of Compositae weeds: EST libraries, microarrays, and evidence of introgression. *American Journal of Botany*, **99**, 209–218.

Lehe R, Hallatschek O, Peliti L (2012) The rate of beneficial mutations surfing on the wave of a range expansion. *PLoS Computational Biology*, **8**, e1002447.

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.

Li L, Stoeckert CJ Jr, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, **13**, 2178–2189.

Linder CR (2000) Adaptive evolution of seed oils in plants: accounting for the biogeographic distribution of saturated and unsaturated fatty acids in seed oils. *The American Naturalist*, **156**, 442–458.

Lockwood JL, Cassey P, Blackburn TM (2009) The more you introduce the more you get: the role of colonization pressure and propagule pressure in invasion ecology. *Diversity and Distributions*, **15**, 904–910.

Löytynoja A, Goldman N (2008) Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*, **320**, 1632–1635.

Lu A, Guindon S (2014) Performance of standard and stochastic branch-site models for detecting positive selection among coding sequences. *Molecular Biology and Evolution*, **31**, 484–495.

Macknight R, Bancroft I, Page T *et al.* (1997) FCA, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. *Cell*, **89**, 737–745.

Markova-Raina P, Petrov D (2011) High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Research*, **21**, 863–874.

Marrs RA, Sforza R, Hufbauer RA (2007) When invasion increases population genetic structure: a study with *Centaurea diffusa*. *Biological Invasions*, **10**, 561–572.

Mayrose M, Kane NC, Mayrose I, Dlugosch KM, Rieseberg LH (2011) Increased growth in sunflower correlates with reduced defences and altered gene expression in response to biotic and abiotic stress. *Molecular Ecology*, **20**, 4683–4694.

Mugford SG, Matthewman CA, Hill L, Kopriva S (2010) Adenosine-5′-phosphosulfate kinase is essential for *Arabidopsis* viability. *FEBS Letters*, **584**, 119–123.

Muth NZ, Pigliucci M (2006) Traits of invasives reconsidered: phenotypic comparisons of introduced invasive and introduced noninvasive plant species within two closely related clades. *American Journal of Botany*, **93**, 188–196.

Nawrath C, Heck S, Parinthawong N, Métraux J (2002) EDS5, an essential component of salicylic acid–dependent signaling for disease resistance in *Arabidopsis*, is a member of the MATE transporter family. *The Plant Cell*, **14**, 275–286.

Ndamukong I, Abdallat AA, Thurow C *et al.* (2007) SA-inducible *Arabidopsis* glutaredoxin interacts with TGA factors and suppresses JA-responsive PDF1.2 transcription. *The Plant Journal: For Cell and Molecular Biology*, **50**, 128–139.

Pagnussat GC, Yu H-J, Ngo QA *et al.* (2005) Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. *Development*, **132**, 603–614.

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013) On the accumulation of deleterious mutations during range expansions. *Molecular Ecology*, **22**, 5972–5982.

Penn O, Privman E, Landan G, Graur D, Pupko T (2010) An alignment confidence score capturing robustness to guide tree uncertainty. *Molecular Biology and Evolution*, **27**, 1759–1767.

Portereiko MF, Sandaklie-Nikolova L, Lloyd A *et al.* (2006) NUCLEAR FUSION DEFECTIVE1 encodes the *Arabidopsis* RPL21M protein and is required for karyogamy during female gametophyte development and fertilization. *Plant Physiology*, **141**, 957–965.

Powles SB, Yu Q (2010) Evolution in action: plants resistant to herbicides. *Annual Review of Plant Biology*, **61**, 317–347.

Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ (2008) Adaptive evolution in invasive species. *Trends in Plant Science*, **13**, 288–294.

Privman E, Penn O, Pupko T (2012) Improving the performance of positive selection inference by filtering unreliable alignment regions. *Molecular biology and Evolution*, **29**, 1–5.

Reeves PH, Coupland G (2001) Analysis of flowering time control in *Arabidopsis* by comparison of double and triple mutants. *Plant Physiology*, **126**, 1085–1091.

Renaut S, Grassa CJ, Moyers BT, Kane NC, Rieseberg LH (2012) The population genomics of sunflowers and genomic determinants of protein evolution revealed by RNAseq. *Biology*, **1**, 575–596.

Ross J, Li Y, Lim E, Bowles DJ (2001) Higher plant glycosyltransferases. *Genome Biology*, **2**, reviews 3004.

Rouhier N, Lemaire SD, Jacquot J-P (2008) The role of glutathione in photosynthetic organisms: emerging functions for glutaredoxins and glutathionylation. *Annual Review of Plant Biology*, **59**, 143–166.

Scaglione D, Lanteri S, Acquadro A *et al.* (2012) Large-scale transcriptome characterization and mass discovery of SNPs in globe artichoke and its related taxa. *Plant Biotechnology Journal*, **10**, 956–969.

Slatkin M, Excoffier L (2012) Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics*, **191**, 171–181.

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Stewart CN, Tranel PJ, Horvath DP *et al.* (2009) Evolution of weediness and invasiveness: charting the course for weed genomics. *Weed Science*, **57**, 451–462.

Sun M (1997) Population genetic structure of yellow starthistle (*Centaurea solstitialis*), a colonizing weed in the western United States. *Canadian Journal of Botany*, **75**, 1470–1478.

Turner KG, Hufbauer RA, Rieseberg LH (2014) Rapid evolution of an invasive weed. *The New Phytologist*, **202**, 309–321.

USDA, NRCS (2014) *The PLANTS Database* (http://plants.usda.gov, 13 September 2014). National Plant Data Team, Greensboro, North Carolina.

Waterhouse RM, Zdobnov EM, Kriventseva EV (2011) Correlating traits of gene retention, sequence divergence, duplicability and essentiality in vertebrates, arthropods, and fungi. *Genome Biology and Evolution*, **3**, 75–86.

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences: CABIOS*, **13**, 555–556.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, **22**, 2472–2479.

K.H. and L.H.R. conceived the study; K.H. and K.T. corralled the data; K.T. classified the species; K.H., D.B., S.H. and M.H. conducted the analyses; K.H. wrote the manuscript and all contributors edited.

Data accessibility

Assembly data are publicly available (see Table S1, Supporting information). Code for several of the analyses performed in this study is available as a GitHub repository (https://github.com/kgturner/InvasionSyndicate).

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Assemblies used in this study.

Table S2 Known target site resistance genes for several common herbicides that were examined in this study.

Table S3 Loci with dN/dS > 1 for pairwise comparisons between native and introduced transcriptomes, their best hit in the TAIR database, the identity of homologous genes among species, and gene ontology terms ($P < 0.05$).

Table S4 Significant loci for branch and branch-site models in PAML for comparisons between native and introduced genotypes and a closely related outgroup.

Table S5 PAML results for three orthogroups that showed evidence of significant changes in dN/dS ratio (ω) in native vs. introduced genotypes.

Table S6 Genes identified as significant from the Fitmodel and PAML site models analyses along with their functional descriptions based on BLASTP to TAIR.

Table S7 Total number of amino acid differences predicted to be deleterious using Provean and the total number of amino acid substitutions that are predicted to be derived deleterious alleles.

Table S8 Total number of alignments tested for each analysis.

Appendix S1 Supplementary methods.