RAPID ADAPTATION OF FLORAL PHENOTYPES IN WEEDY RADISH,
*R.R. RAPHANISTRUM*


By

Amanda Charbonneau

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Genetics - Doctor of Philosophy

2018

**ABSTRACT**

RAPID ADAPTATION OF FLORAL PHENOTYPES IN WEEDY RADISH,
*R.R. RAPHANISTRUM*

By

Amanda Charbonneau

Agricultural weeds cause billions of dollars' worth of damage worldwide as well as reducing

yields, however we often know very little about where they came from, or how they adapt to

farming techniques. Agricultural fields are human created environments quite unlike anything in

nature and are relatively harsh environments that can exert strong selective pressures. Yearly

tilling, for example, likely selects for plants that both quickly reproduce, and can survive in

disturbed soils. While some plants with generalist phenotypes might be well suited for thriving in

these conditions, others, like *Raphanus raphanistrum ssp. raphanistrum* (weedy radish), have

likely evolved to become weeds. To better understand how these agricultural weeds evolve, I

have phenotypically and genotypically characterized weedy *R.r. raphanistrum* and it's close

relatives.

In chapters one and two, I show that weedy *R.r. raphanistrum* is most closely related to native

populations of *R.r. raphanistrum*, but that these two ecotypes have very different flowering

phenotypes. Weedy *R.r. raphanistrum* flowers in approximately thirty days, while the native

plants take fifty to one hundred and fifty days to flower. This demonstrates a likely adaptation to

agriculture, and in chapter two I find several loci that may contribute to these phenotypic

differences. In chapter three, I analyze differential expression patterns in two selection lines

derived from weedy *R.r. raphanistrum*, to determine genes that underlie differences in floral

morphology. These genes should contribute to differences in anther exsertion, which in turn controls how pollen is dispersed onto pollinators.

Together, these studies answer basic questions about how evolution works on a short time scale and provide insights into the adaptations of one of the world's most damaging agricultural weeds. More broadly, these studies demonstrate that weedy radish is a good system for studying rapid evolution in response to both natural and artificial selection and lay the groundwork for future work. In particular, chapters one and two will be useful for broad comparisons across agricultural weeds to determine whether weeds use similar strategies for invading croplands, which would tell us not only about the repeatability of evolution, but also be potentially useful in reducing agricultural losses due to weeds.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

**Chapter I: Weed evolution: Genetic differentiation among wild, weedy, and crop radish**

The work described in this chapter is in review as the following manuscript:

Charbonneau, A., Tack, D., Lale, A., Goldston, J., Caple, M., Conner, E., Barazani, O., Ziffer-Berger, J., Dworkin, I., Conner, J. (2018). Weed evolution: Genetic differentiation among wild, weedy, and crop radish. Evol. Apps.

# Abstract

Approximately 200 weed species are responsible for more than 90% of crop losses and these comprise less than one percent of all named plant species, suggesting that there are only a few evolutionary routes that lead to weediness. Agricultural weeds can evolve along three main paths: they can be escaped crops, wild species, or crop-wild hybrids. We tested these three hypotheses in weedy radish, a weed of small grains and an emerging model for investigating the evolution of agricultural weeds. We grew over 2400 individuals from 43 populations representing all major species and sub-species in the radish genus *Raphanus* in a series of common garden experiments, to estimate genetic differentiation in flowering time. We also assayed genetic relationships of 34 of these populations using both STRUCTURE and SmartPCA. Our findings suggest that the agricultural weed radish *R.r. raphanistrum* is most genetically similar to native populations of *R.r. raphanistrum* and is likely not a feral crop or crop hybrid. We also show that weedy radish flowers more rapidly than any other *Raphanus* population or cultivar, which is consistent with rapid adaptation to the frequent and severe disturbance that characterizes agricultural fields.

2

## Introduction

Modern farming provides a nearly ideal environment for fostering the rapid evolution of weeds. Plants growing in novel environments often evolve rapidly (Buswell et al., 2011), and human alterations of the environment often impose strong selection (Palumbi, 2001; Reznick and Ghalambor, 2001). In invasive species, rapid adaptation is also correlated with high levels of standing genetic variation, repeated introductions, and high levels of environmental disturbance (Sakai et al., 2001). Agricultural fields combine all of these factors. They are a created ecosystem unlike anything found in nature, where humans impose frequent and regular disturbance from tilling and harvesting, and contaminated seed stocks can repeatedly introduce large numbers of weed seeds. Understanding both the origins and adaptations of weeds could guide improvements in weed management (McNeill, 1976; Müller-Schärer et al., 2004; Neve et al., 2009; Stewart Jr et al., 2009; Vigueira et al., 2012) as well as provide insight into approaches for preventing the evolution of future weeds (Higgins et al., 1978).

There are three potential origin routes for agricultural weeds: crops going feral, wild populations invading fields, or crop-wild hybridization (de Wet, 1966; de Wet and Harlan, 1975; Vigueira et al., 2012); and these differing histories may have detectably different phenotypic and genetic effects. Escaped crops, for example, may already be resistant to herbicides or able to survive in disturbed habitats (Vigueira et al., 2012). These different origins would also leave a distinct genetic signature at neutral markers. Weeds could have a wild origin, with either a wild population that was pre-adapted, or one that rapidly evolved to weediness. Here we would expect strong genetic resemblance between the weed and one or more native populations (Vigueira et al., 2012). Alternatively, the weeds could be crop-wild hybrids, in which case we would expect the weeds to be a genetic mixture of cultivars and native plants with relatively high genetic

3

diversity compared to the parental crop. Finally, if the weeds are most genetically similar to the cultivars at neutral markers, this would suggest that they descend from escaped crops. As feral crops would suffer from the founder effect of escaping cultivation after having undergone artificial selection, we might also expect weeds derived from crops to be less genetically variable than either wild invaders or hybrids (Vigueira et al., 2012).

Although these potential weed origins are widely discussed in the literature (Baker, 1974, 1991; de Wet, 1966; Dekker, 2011; Ellstrand et al., 2010, 2013; Gressel, 2005; Harlan, 1992b; Vigueira et al., 2012), experimental evidence for most species is lacking. In the handful of modern studies which have determined the species or population of origin for a given weed, most have found it to be a feral crop, or crop-hybrid (Desplanque et al., 1999; Ellstrand et al., 2010; Kanapeckas et al., 2016; Muller et al., 2010; Zizumbo-Villarreal et al., 2005, but see: Lai et al., 2008; Pickersgill, 1971; Yoon et al., 2007). However, this represents a lack of data combined with bias, as many examples come from the Ellstrand et al (2010) review, which focused only on finding weeds that had a crop origin. Studies of weed origins can be challenging, as the genetic similarity of populations in weed-crop-wild triads, especially when combined with subsequent hybridization, can make it difficult to identify the true source population for the weed (Ellstrand et al., 2010; Vigueira et al., 2012).

Agricultural weed adaptations have also been relatively neglected, with the notable exception of adaptations to human control, particularly the evolution of crop mimicry and herbicide resistance as reviewed in (Barrett, 1983; Neve et al., 2009; Vigueira et al., 2012). However, weeds must already be well-adapted to agricultural habitats before they are problematic enough for human control practices to be implemented. Some attention has been paid to the key adaptations of seed dormancy and shattering in the evolution of agricultural weeds from crop ancestors (Ellstrand et

al., 2010; Vigueira et al., 2012), but there has been little work on the evolution of a shortened life-cycle. In an agricultural setting, frequent and regular disturbances from plowing and harvesting likely exerts a strong selection on weeds for rapid flowering and seed set (Barrett, 1983; Warwick and Stewart, 2005).

Weedy radish, *Raphanus raphanistrum ssp. raphanistrum*, is an emerging model for studying both rapid adaptation and weed evolution (Campbell et al., 2006, 2009b; Conner et al., 2011, 2014; Klinger et al., 1991; Ridley and Ellstrand, 2008; Sahli et al., 2008; Snow and Campbell, 2005), and is a member of one of the four major weed and crop families (*Brassicaceae*). Determining the origins of weedy radish is tractable because the genus *Raphanus* includes only three named species. The genus likely originated in the Mediterranean, as native populations exist only there (I. Al-Shehbaz, pers. comm.), and all members of the genus are self-incompatible. *R. raphanistrum* is divided into two subspecies, with *R. r. raphanistrum* including native Mediterranean populations as well as the globally-distributed weed, and *R. r. landra* existing only as native Mediterranean populations. The crop *R. sativus* is divided into four major types two root crops (European radishes and Asian daikon) and two fruit crops (Oilseed and edible-pod Rattail). *R. pugioniformis* is a little-studied endemic of the eastern Mediterranean (Ziffer-Berger et al., 2014). The relationships among these species are not well-resolved (Ziffer-Berger et al., 2014), but an analysis of cDNA sequence from one population of each of eight *Raphanus* taxa (not including *R. pugioniformis*) provided strong support for the monophyly of the crop cultivars, as well as monophyly of native and weedy *R.r. raphanistrum* (Shen et al., 2013).

While native *Raphanus* is found only near the Mediterranean, weedy populations are found on every continent except Antarctica (Holm, 1997). Weedy *R.r. raphanistrum* is a common

contaminant of small grain fields and is considered one of the world's worst agricultural weeds (e.g., Blackshaw et al., 2002; Culpepper, 2012; Holm, 1997; Schroeder, 1989; Warwick and Francis, 2005; Webster and Macdonald, 2001). Weedy radish has been used extensively in ecological and evolutionary studies, particularly for plant-insect interactions (Agrawal et al., 2002; Bett and Lydiate, 2003; Conner, 2002; Conner et al., 2003; Devlin and Ellstrand, 1990; Irwin et al., 2003; Lehtilä and Strauss, 1999; Malik, 2009; Mazer and Schick, 1991; Morgan and Conner, 2001; Snow et al., 2001; Stanton et al., 1986), and has genomic resources (Moghe et al., 2014) that make it an ideal study system to answer questions about both the origins and adaptations of agricultural weeds. Studies have shown that weedy radish lacks a signal of isolation by distance (Barnaud et al., 2013a,b; Kercher and Conner, 1996), likely due to human-mediated movement of large numbers of seeds long distances as a grain contaminant.

Previous work has shown that eight populations of weedy radish flowered much more quickly than one native *R. r. raphanistrum* population in a greenhouse common garden (Sahli et al., 2008). Similarly, five weedy radish populations flowered faster than five root-crop radish cultivars in greenhouse (Hegde et al., 2006) and field (Ridley and Ellstrand, 2008) common gardens. However, since flowering time has only been reported for one native population, and the phylogeny in Shen et al (2013) was based on only a single population of each *Raphanus* taxon, it is not clear whether the weeds have evolved earlier flowering as an adaptation to the frequent and regular disturbance in agricultural fields. This study aims to determine both the genetic relationships between crop, native and weedy *Raphanus* populations, and the extent to which these weeds have evolved earlier flowering time as an adaptation to agricultural fields. We estimated genetic differentiation for flowering time and molecular markers for all named species

and subspecies in the genus *Raphanus*, including a total of 23 wild populations and 21 crop cultivars from all four groups (Table 1-1).

**Table 1-1: Summary of the eight flowering time common garden experiments.** Experiments took place at one of two sites in Michigan over a period of 11 years. NPops is number of populations in each experiment. Number of individuals in each population is given as NperPop. TotalN is the total number of individuals in that experiment

| Experiment | Year | Location | FieldGH | NPops | NperPop | TotalN | References |
|---|---|---|---|---|---|---|---|
| G-03 | 2003 | KBS | GH | 9 | 22-46 | 306 | Parentals (Sahli et al., 2008) |
| G-04 | 2004 | KBS | GH | 9 | 58-142 | 877 | Offspring (Sahli et al., 2008) |
| F-05 | 2005 | KBS | Field | 6 | 64-88 | 442 | Offspring (Sahli et al., 2008) |
| G-10 | 2010 | KBS | GH | 4 | 8-22 | 55 | |
| F-12 | 2012 | KBS | Field | 13 | 7-10 | 127 | |
| F-13 | 2013 | MSU | Field | 23 | 10 | 229 | |
| G1-13 | 2013 | KBS | GH | 15 | 10 | 150 | |
| G2-13 | 2013 | KBS | GH | 9 | 14-30 | 254 | |

## Materials and Methods

**Populations**

Eleven *R. r. raphanistrum* populations from the native Mediterranean range were included, with 6 from the western part of the range (collected in Spain or France) and 5 eastern (collected in Israel). We used eight populations from outside the native range, these were collected as weeds

of agricultural fields or disturbed areas in USA, Europe, and Australia. Three populations of *R. r. landra* from Spain and one *R. pugioniformis* population from Israel were also included, as were twenty-one crop varieties purchased from seed companies (Table S1). The natural populations were collected by a variety of individuals using a variety of methods, but in all cases the goal was to sample the genetic variation of each population in an unbiased manner.

**Genotyping**

To assay patterns of neutral genetic differentiation among these populations, we used a panel of 13 CAPS (cleaved amplified polymorphic sequences) in addition to the 8 SSR (microsatellite) markers from Sahli et al. (2008). To create the CAPS markers, cDNA sequencing of seven lines of *Raphanus* populations and cultivars Moghe et al. (2014) was used to assemble and align line-specific contigs against each other (Fig. A1-2). Genotyping was completed on 8-10 randomly-sampled plants from each of 34 populations for a total of 338 individuals (Table S1); some of the non-native and crop cultivars were left out of the genotyping to improve balance across the different groups and save genotyping costs. Two microsatellites, Na10H06 and Na14E08, had fairly high numbers of missing genotypes (26% and 31% respectively); however, these were concentrated among crop cultivars, and dropping these markers did not have a qualitative impact on our results, so the data presented here includes all markers. Standard summary statistics were computed using the **R** packages 'adegenet' (v.2.0.1) and 'pegas' (v.0.10).

**SmartPCA**

SmartPCA (v.13050 – from the program Eigensoft 6.0.1) (Patterson et al., 2006) was used to perform an eigendecomposition optimized for genomic data to rotate the data onto a set of orthogonal axes defined by the amount of variation explained. Although this method rotates,

rather than clusters the data, it can reveal hidden data structure. SmartPCA was originally developed for datasets where the number of markers vastly exceeds the number of individuals genotyped, where performing a standard PCA would be difficult. Our dataset does not fit this expectation, as we have relatively few markers compared to individuals, however SmartPCA still outperformed standard PCA. The SmartPCA algorithm tolerates missing genotypes, so we could use all individuals and markers in our analysis rather than dropping any with missing data. A standard PCA gave qualitatively similar result, just with fewer usable data points (Fig. A1-3).

As the SmartPCA algorithm is designed for biallelic markers, each SSR marker was expanded into several biallelic markers as described in Patterson et al. (2006), prior to analysis using a custom script. Experimentally determined linkage groups were used as a proxy for chromosomes http://radish.plantbiology.msu.edu/. Markers that could not be assigned to linkage groups were given unique chromosome numbers. The total number of linkage groups (7) plus singletons (3) was used as the chromosome number (Table S2); note that this is one more than the nine *Raphanus* chromosomes.

**STRUCTURE**

Our second approach – STRUCTURE (Pritchard and Stephens, 2000) – clusters individuals by their genetic similarity. We used the command line version (v2.3.4), with a burn in of 500,000 cycles followed by one million Markov chain Monte Carlo iterations. We used the correlated allele frequencies (or F) model, which computes values similar to Fst to model genetic differentiation and does not require genetic linkage information. Our populations were not constrained to a single Fst, and alpha (the admixture parameter) was inferred for the overall

dataset as suggested in the manual (v. 2.3, Pritchard et al., 2010). We used the default settings for all priors.

STRUCTURE has frequently been used to estimate 'K', the number of distinct genetic groups present in a dataset (Evanno et al., 2005; Falush et al., 2016; Puechmaille, 2016). Although STRUCTURE was not *designed* to determine K, comparing the output of analysis at multiple values of K can suggest which values of K are most likely. For each potential value of K from 3 to 24, we ran 20 separate iterations of the program, each with a different randomization of the data input order. From this, we calculated optimal values for K using both the mean log probability of K for each iteration as in Pritchard and Stephens (2000) and the largest delta K as described in Evanno et al. (2005). These calculations were run using the web-based version of Structure Harvester (Web v0.6.94 July 2014, Plot vA.1 November 2012, Core vA.2 July 2014) (Earl and vonHoldt, 2012).

Both the mean log-likelihood and delta log-likelihood method (Evanno et al., 2005) predict a K=19, however this is likely spuriously high. At K=19, 85% of our replicates have one or more 'ghost' (Puechmaille, 2016) or 'spurious' (Guillot et al., 2005) clusters (data not shown), which may represent admixture from unsampled, ancestral populations, or an analytical artifact (Guillot et al., 2005; Puechmaille, 2016). The number of these clusters decreases with decreasing K, with ghost clusters disappearing below K=10, and spurious clusters below K=9. However, major population groupings are robust over the entire range of K values, and across multiple runs of the same K (data not shown). Therefore, we used K=7 based on *a priori* groupings allowing for divergence between the three major wild taxa: *R. pugioniformis*, *R.r. landra*, and *R.r. raphanistrum*, and the four major crop groups; as several studies have suggested have separate

origins for the two root crops (Kong et al., 2011; Lü et al., 2008; Muminović et al., 2005; Wang et al., 2008; Yamagishi and Terachi, 2003; Yamane et al., 2005, 2009).

**AMOVA**

We used an analysis of molecular variance (AMOVA) to test three hypotheses about the origin of weedy radish. In all cases, we used raw pairwise distances, set 'filter' to TRUE, did not calculate individual variance by haplotype, and ignored missing data, as removing missing values had no qualitative effect on the results. To test whether weeds are likely to be feral crops, we analyzed a subset of the data that included only *R. sativus* and non-native *R.r. raphanistrum.* In this model, population of origin was nested in group, where group was either crop (*R. sativus*) or weed (non-native *R.r raphanistrum*).

To test whether weeds are likely derived from native *R.r. raphanistrum*, we analyzed a subset of the data that included only native and non-native *R.r. raphanistrum*, where population of origin was nested in group- either native or non-native.

 To test whether weeds are more closely related to native *R.r. landra*, we analyzed a subset of the data that included only *R.r. landra* and non-native *R.r. raphanistrum,* where population of origin was nested in group (sub-species). We obtained p values for all tests using a permutation test with 500 repetitions.

All AMOVA were run using the ade4 reimplementation from the 'poppr' package (v2.6.1, Kamvar et al., 2014) for R (v 3.2.2, R Core Team, 2015). Graphs were produced using custom scripts and ColorBrewer (v.1.1-2, Neuwirth, 2014). All SmartPCA and STRUCTURE analyses were run on the HPCC at Michigan State University. We plotted results from SmartPCA and STRUCTURE analysis using custom bash and **R** scripts (v 3.2.2, R Core Team, 2015).

**Phenotyping**

To test for genetic differentiation in flowering time, we combined data from eight common garden experiments performed over a period of 11 years and including a total of 2441 plants (Table 1-1). Five of these experiments (G-03, G-04, F-05, G-10, and G2-13), had a relatively large number of individuals per population (mean = 49), but a small number of populations per experiment (range 4-9), and while there was overlap between experiments, each had a unique combination of populations. To complement results from these trials, we also included data from three additional common garden experiments (two field, one greenhouse) with both more populations each, and more overlap between populations (F-12, F-13, and G1-13). Two experiments (G-04, F-05) used seeds from full-sibling families as described in Sahli et al. (2005); we did not account for this in the analysis due to the complexity of the models used (see below) and the fact that the other experiments also include natural and unknown family structure.

In all experiments, locations of individual plants were randomized with respect to population, and seeds were removed from pods prior to planting to minimize variability in germination times. All plants were monitored daily for germination and until they flowered or died without flowering. In the field experiments, plants were left in the field to over-winter and monitored for flowering the following spring. In the greenhouse experiments, 225 *R. raphanistrum* plants from four of the native populations (CBES, MAES, PBFR, and SAES) did not flower in 82-120 days, and these received a vernalization treatment. We recorded germination date and date of first flower on all plants; the difference between these dates is our measure of flowering time. Plants were watered as needed.

**Analysis**

Phenotypic data from all eight experiments were concatenated for analysis as a single dataset. Only a small subset of plants from four native populations were subjected to vernalization (see above). Similarly, although most populations are represented in multiple experiments, the datasets are not balanced. However, if we examine flowering time in the three populations that were included in four or five experiments including both greenhouse and field, we see that there is much more variance in flowering time among populations than among experiments within populations. Still, we accounted for these aspects of the dataset by analyzing the flowering time data using two mixed models in a Bayesian framework, similar to a modern meta-analysis. The first model includes only *R.r. raphanistrum* plants, that is, those with little or no vernalization requirement. The second includes only native *Raphanus*, which have variable vernalization requirements. To determine whether vernalization time had an effect on parameter estimates, we also ran both models with and without vernalization ($v$) as a fixed effect. In all plots of modeled values, both estimates are shown. This method resulted in two reasonably balanced datasets and provides two estimates for native *R.r. raphanistrum*. We also provide un-modeled population level estimates of both flowering time and vernalization requirement.

We modeled days to flower as a function of geographic origin to verify the differences in flowering time between native and non-native *R.r. raphanistrum* populations reported by Sahli et al. (2008), but with multiple native populations. We also tested for differences between native populations from eastern and western Mediterranean. For this analysis, we used a subset of the data that only included *R.r. raphanistrum* populations (Table S1) and controlled for a number of covariates.

13

This model took the form of:

$$f_i \sim N\big(\beta_0 + \beta_r \chi_{1,i} + \beta_m \chi_{2,i} + \beta_{DOYGS} \chi_{3,i} + \beta_v \chi_{4,i} + \eta_{j[i]} + \eta_{k[i]}, \sigma_f^2\big)$$

$$with$$

$$\eta_j \sim N\big(0, \sigma_{experiment}^2\big) \qquad (1)$$

$$and$$

$$\eta_k \sim N\big(0, \sigma_{population}^2\big)$$

where $f$ is the number of whole days from germination to flowering, if available, the number of

whole days from germination to bolting for plants that bolted but didn't flower, and the number

of whole days from germination until the last day of monitoring in the fall, plus one, for plants

that survived without flowering until this date but did not survive the winter. $r$ is the geographic

region of origin (western or eastern Mediterranean, or outside this native range). $m$ is a factor

representing whether the seed producing that individual was collected directly from the

population of origin or generated in the greenhouse. $DOYGS$ is the day of year that the plant

germinated (1-365) scaled such that 1 is the spring equinox in the northern hemisphere. $v$ is the

number of days each plant was vernalized, if any. The random effect $\eta_j$ is the experiment that

each plant was in, and accounts for year-to-year variation as well as field vs. greenhouse and

other variation in experimental protocols. The random effect $\eta_k$ accounts for variance among

populations within geographic regions.

To look for differences in flowering time among *Raphanus* species and sub-species from within

the native range, we used a subset of the data that included only populations collected from

within the Mediterranean region. This model took the form of:

14

$$f_i \sim N\left(\beta_0 + \beta_u \chi_{1,i} + \beta_m \chi_{2,i} + \beta_{DOYGS} \chi_{3,i} + \beta_v \chi_{4,i} + \eta_{j[i]} + \eta_{k[i]}, \sigma_f^2\right)$$

$$with$$

$$\eta_j \sim N\left(0, \sigma_{experiment}^2\right) \qquad (2)$$

$$and$$

$$\eta_k \sim N\left(0, \sigma_{population}^2\right)$$

where $u$ is a factor representing species or sub-species designation (*R.r. raphanistrum, R.r. landra*, and *R. pugioniformis*), and the other terms are as in the previous model.

All phenotypic models were run using the 'MCMCglmm' package (v.2.2, Hadfield, 2010) for **R** (v 3.2.2, R Core Team, 2015). Graphs were produced using custom scripts and ColorBrewer, (v.1.1-2, Neuwirth, 2014).

Scripts and all data required for all analyses are currently bundled with the supplement, but on publication will be available via the authors github page.

## Results

**Distinct Clustering of wild and weedy populations of *Raphanus* species and sub-species**

The first two principal components from SmartPCA show clear differentiation between the three named species, i.e. *R. raphanistrum, R. sativus,* and *R. pugioniformis* (Fig. 1-1). In addition to these groupings, we also observe differentiation between *R.r. raphanistrum* and the other sub-species *R.r. landra*. Although the *R. raphanistrum* sub-species form a continuous arc in the space

defined by the first two PCs, we find separation between most native and non-native *R.r. raphanistrum* populations.

Populations taken from within the Mediterranean region are relatively tightly clustered at either end of the arc (*R.r. landra* in pink, native-range *R.r. raphanistrum* in blue/purple), while populations from outside the region form a more diffuse (orange) group in the center. Although non-Mediterranean *R.r. raphanistrum* populations were taken from sites spread across three continents, they occupy a similar area in PC space. While these non-natives occupy the same genetic sub-space, individuals from all five populations can be found spread evenly throughout that portion of the plot. Our results using STRUCTURE are consistent with the major groupings from SmartPCA but suggest additional subdivisions (Fig. 1-2). Consistent with the SmartPCA results, *R. pugionformis* is intermediate between daikon and Eastern (Israeli) populations. Native range *R.r. landra* cluster together and are distinct from *R.r. raphanistrum*. Native and non-native range *R.r. raphanistrum* are also clearly distinguished except for some allele sharing between Western natives and the Australian populations (purple color). In addition, STRUCTURE clearly
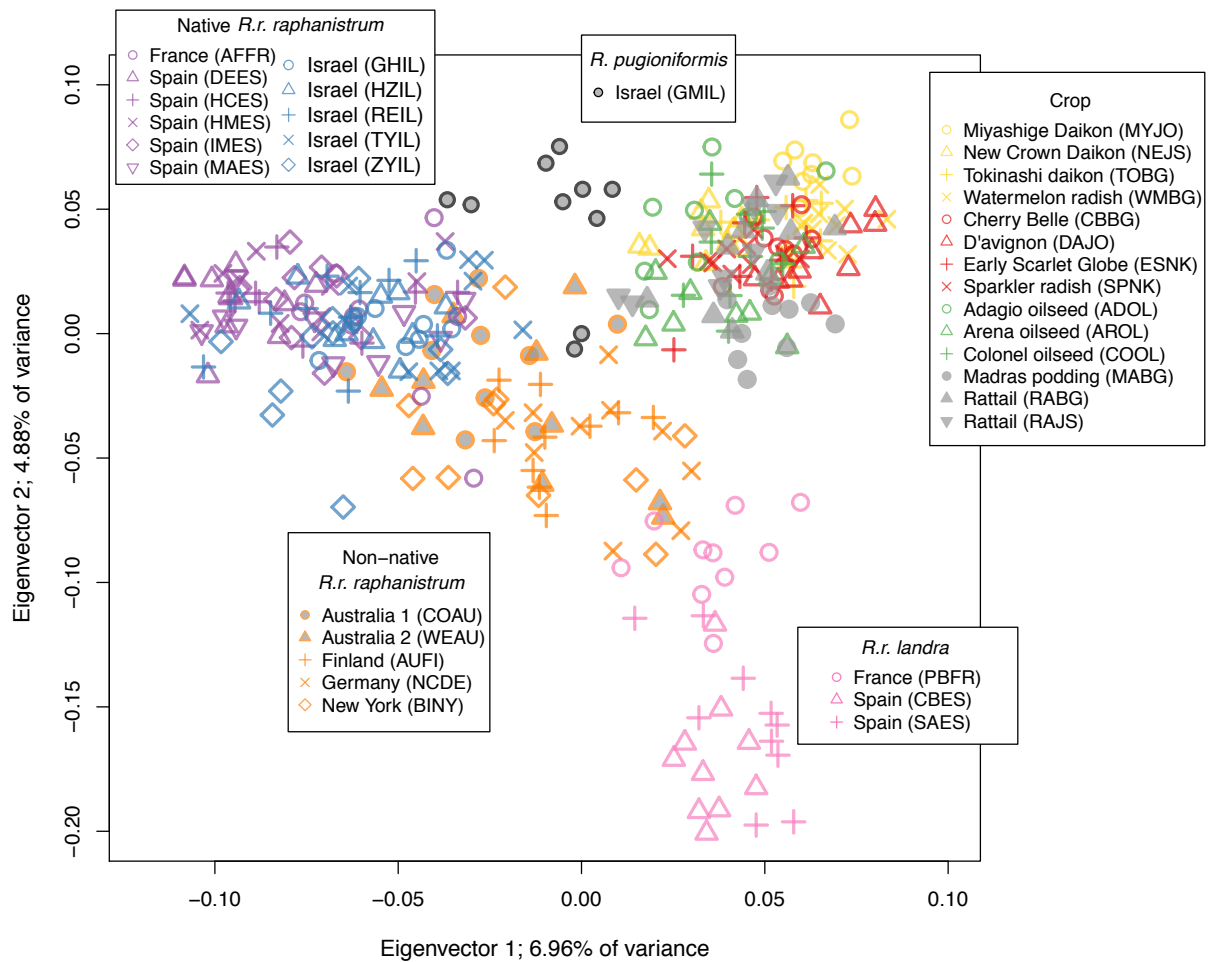
**Figure 1-1: Smart PCA plot of the first two eigenvectors of a principal components analysis of 34 *Raphanus* populations genotyped at presumed neutral markers.** Each point is an individual, and each population is represented by 8-10 individuals. Populations are identified by plotting character and colored to match the putative groups from STRUCTURE (Fig. 1-2) Populations that were mostly admixed in STRUCTURE (GMIL, MABG, RABG, RACA, COAU and WEAU) are shown with gray fill. (See Fig. A1-1. for population abbreviations).

separates the Western and Eastern native populations unlike SmartPCA. Interestingly, although

SmartPCA shows no obvious population-level clustering within the non-native range *R.r.*

*raphanistrum*, STRUCTURE groups the European and North American weeds together, and

suggests that the Australian populations are the product of admixture between this group and the

Western natives.

Pairwise Fst results (Fig. 1-3) are generally consistent with the taxonomy and the SmartPCA and STRUCTURE results, but there are some differences. The four species and subspecies cluster together, except that *R. pugioniformis* is grouped with the crops, and weedy and native *R.r. raphanistrum* form distinct groups. Also consistent are results of AMOVA (Table 1-2), which shows that weedy *R.r. raphanistrum* is significantly different from native *R.r. raphanistrum, R.r. landra* and *R. sativus*. The percent covariance explained by the comparisons between weedy radish and each of these other possible progenitors is least with native *R.r. raphanistrum* (10.1%) and greatest with *R. landra* (19.5%).
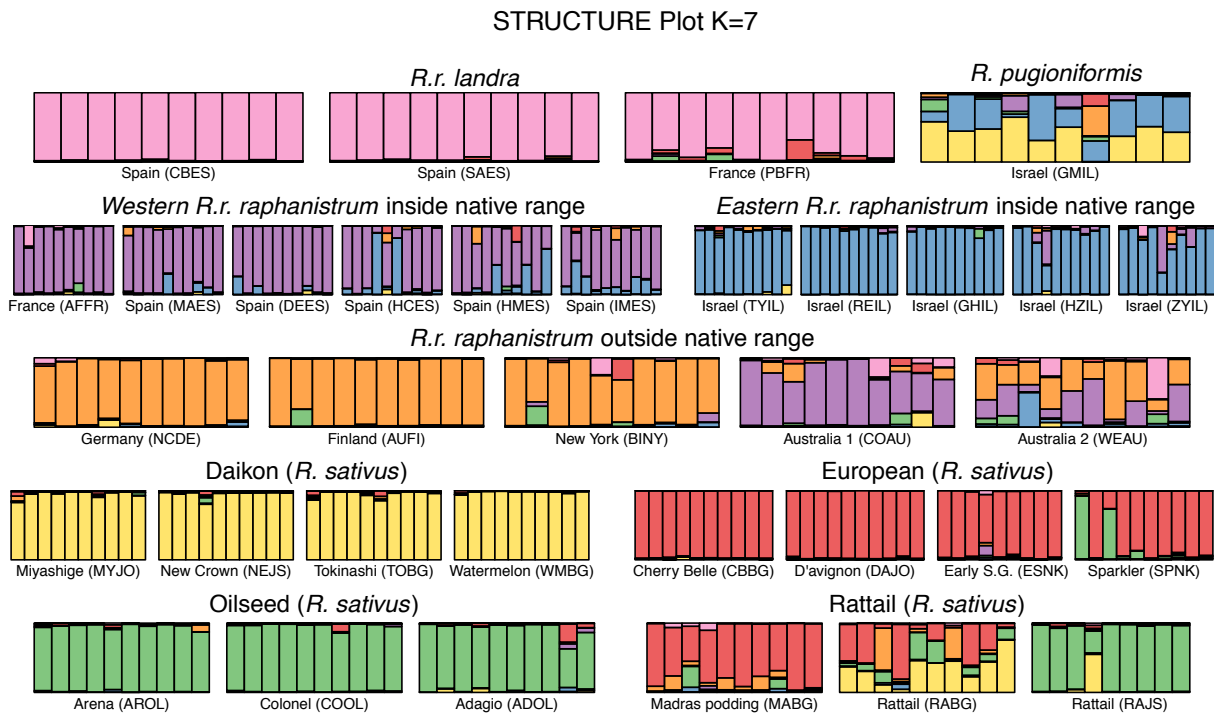


**Figure 1-2: Output from STRUCTURE analysis of 34 *Raphanus* populations genotyped at presumed neutral markers using the admixture model.** Each large rectangle represents a group of 8-10 individuals, with each individual represented by a vertical bar. Populations are grouped into the same putative ecotypes as highlighted in Fig. 1-1. A vertical bar of a single color denotes an individual whose genotypes can be entirely attributed to a single background shared by every individual with that color. Individuals with bars of more than one color show evidence for admixture, with the proportion of each genetic background plotted indicated by proportion of each color.

**Crop varieties are genetically distinct from both natives and weeds**

Both STRUCTURE and SmartPCA clearly separate *R. sativus* from the other species, with the possible exception of some shared daikon and *R. pugioniformis* ancestry noted above. In addition, STRUCTURE (Fig. 1-2) suggests that the daikon, European, and oilseed crops are genetically distinguishable from each other, but that rattails share similarities with each of these other three crops. SmartPCA also suggests some separation between daikons and the other
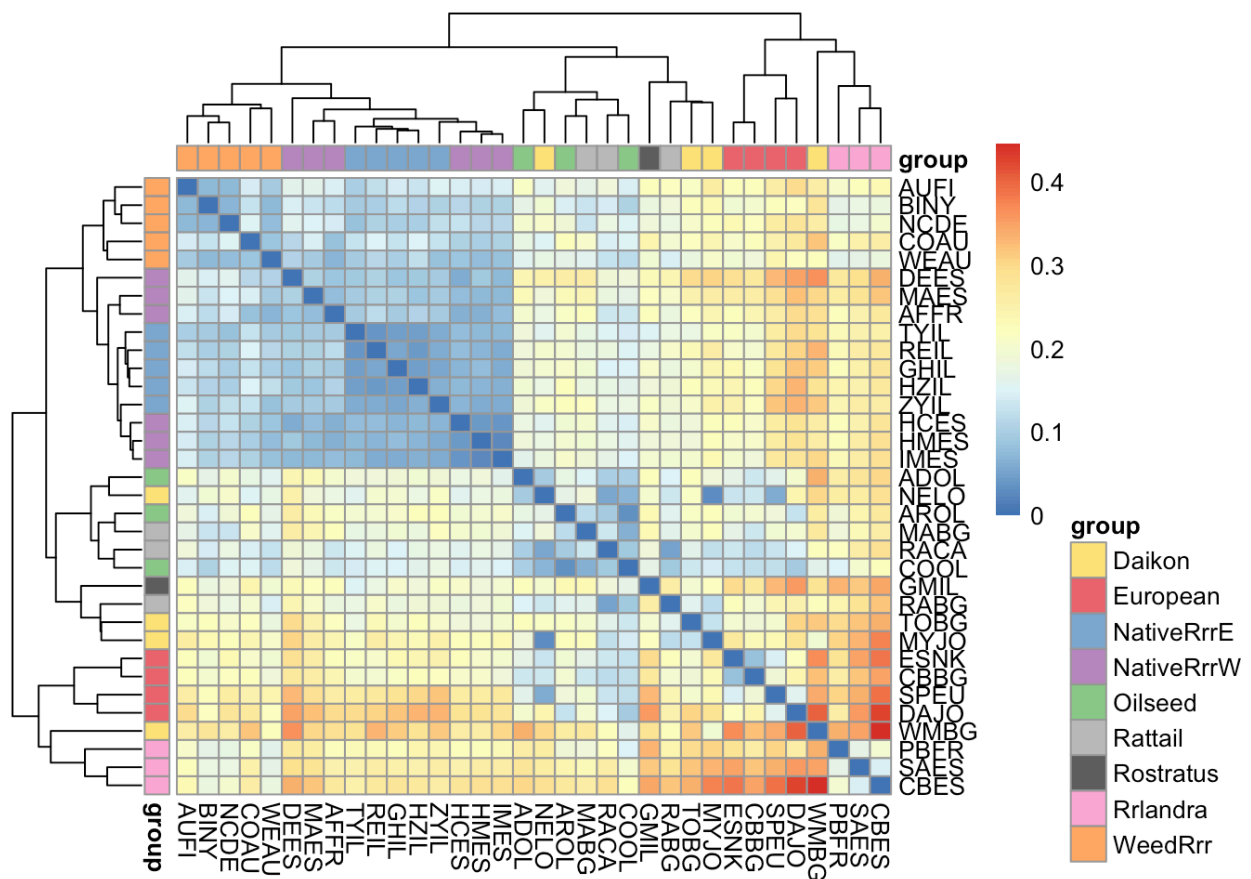


**Figure 1-3: Pairwise Fst calculated for all 21 markers, and clustered by Euclidean distance.** Populations are colored along the axes to match the putative groups from the SmartPCA (Fig 1-1.) and STRUCTURE (Fig 1-2) analyses; for population codes on the other axes, see Table S1.

cultivars (Fig. 1-1). Interestingly, clustering by pairwise Fst suggests that the European crops are more genetically similar to *R.r. landra* than to the rest of *R. sativus* and that *R.r. landra* is relatively distant from the other *R. raphanistrum* sub-species (Fig. 1-3).

**Non-native range *R.r. raphanistrum* flower more rapidly**

In agreement with previous reports, we found that the non-native *R.r. raphanistrum* flower faster than any of the crops, (Hegde et al., 2006; Ridley and Ellstrand, 2008) or natives (Sahli et al., 2008) (Figs. 1-4 and A1-4). In model 1, non-natives flowered about fifty-eight days earlier than the Western range *R.r. raphanistrum*, and about twenty-four days earlier than the Eastern populations (Figs. 1-4 and A1-4). This flowering time difference between the eastern and western populations is consistent with the STRUCTURE analysis that clustered these groups separately (Fig. 1-2). The native *R.r. raphanistrum* population with the fastest flowering time (AFFR) was the only one collected from an active agricultural field, while the slowest flowering (DEES) was the only one collected from an undisturbed habitat (Table S1). *R.r. landra* populations took even longer to flower. In model 2, they required an additional 76 to 139 days to flower over the average native *R.r. raphanistrum* plant, depending on whether vernalization time was included as a fixed effect (Fig. A1-5). Not surprisingly, on average the root crops (daikon and European) flowered more slowly than the crops that are used for their fruits (oilseed and rattail), as flowering causes resources stored in the roots to be reallocated to the fruits.
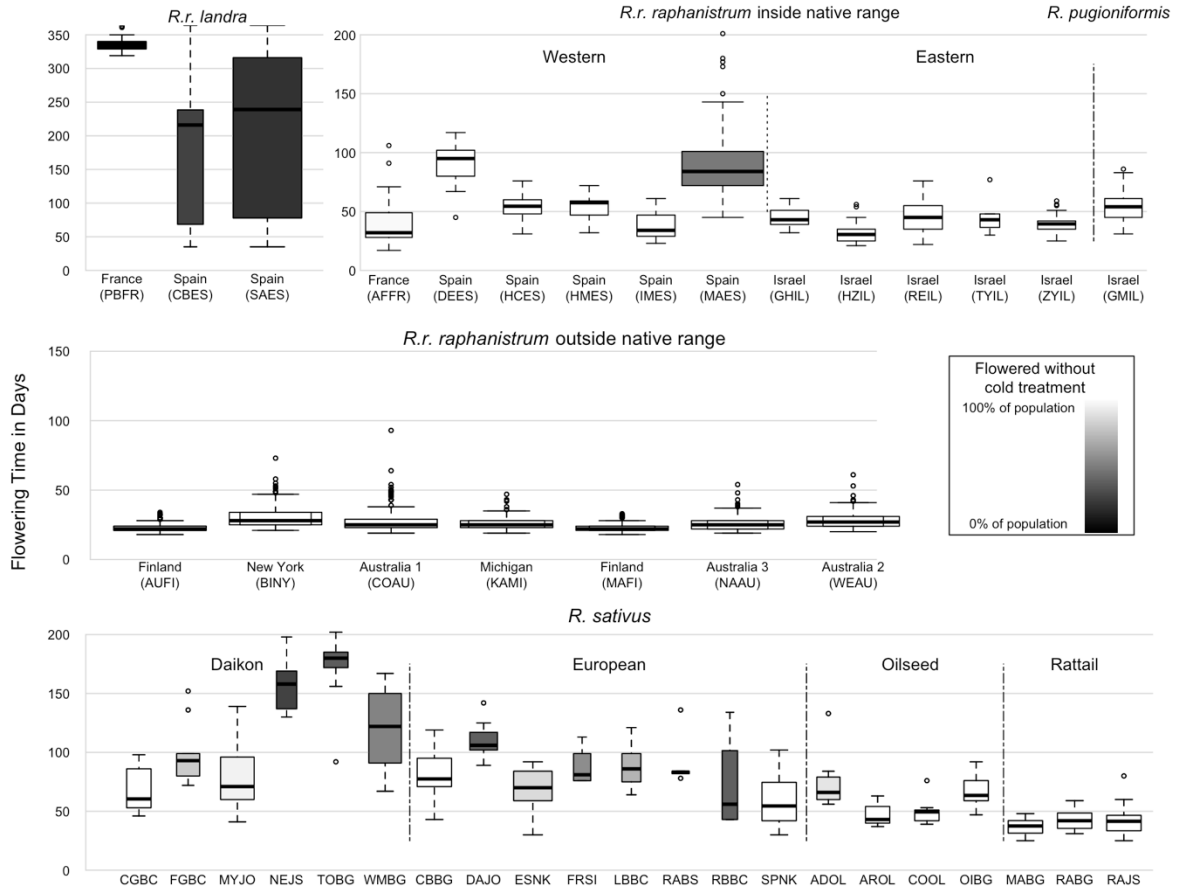
**Figure 1-4: Raw Flowering times in *Raphanus*.** Medians, quartiles, and outliers for raw days from germination to first flower for each population are shown, with shading to denote the proportion of plants that flowered without experiencing vernalization. Boxplot widths are a function of number of individuals per population, with wider plots indicating more individuals. Maximum: AUFI (N=250); minimum: RBBC (N=3), total=2054. Note that scales vary by ecotype, and are very different for *R.r. landra*, which has much longer flowering times than the other taxa. None of *R.r. raphanistrum* populations from outside the native range required vernalization.

There was considerable variation in flowering times among the *R.r landra* and *R. sativus* populations (Figs. 1-4, A1-5, and A1-6). For example, although the *R.r. landra* populations cluster by genetic markers, the French population was slower to flower (due to an absolute vernalization requirement) compared to populations collected in Spain (Figs. 1-4 and A1-5); and the latter had much higher within-population variation in flowering time than any of the other populations studied. In contrast, flowering time estimates for the weeds are consistently rapid

with little within-population variation (Figs. 1-4 and A1-6). Although this study spanned several years and experiments, most of the within-population variation we observed is biological. Any given population has very similar flowering time distributions regardless of the time or location of the experiment, and some native populations consistently have bi-modal distributions (Fig. A1-8)

## Discussion

We assayed a diverse set of populations across the genus *Raphanus*, to address two interconnected questions concerning the evolution of weedy radish. First, what is the likely origin of the weeds; and second, has weedy radish evolved rapid flowering in comparison to its progenitors?

### Origins of radish as an agricultural weed

There are three main pathways to agricultural weediness; feral crops, wild invaders and hybridization, either wild-wild or wild-crop. In cases where the origin of agricultural weeds are known, researchers have frequently found them to be escaped crops (Ellstrand et al., 2010; Vigueira et al., 2012); however, our results do not support this. STRUCTURE did not find evidence for shared genetic background between the non-native *R.r. raphanistrum* and any crop population for values of K ranging from K=3 to K=24 (data not shown), and the crops are significantly different from the weeds by AMOVA F1(<.001) (Table 1-2). These results are inconsistent with the crop origin theory, but are consistent with previous work that found no genetic overlap between crop and weed populations for chloroplast haplotypes (Ridley et al., 2008). Additionally, weeds resulting from de-domestication are expected to have very low genetic diversity (Vigueira et al., 2012), but non-natives in our study have the highest expected

heterozygosity of any group (Table S9). This does not seem to be due to pooling genetically differentiated populations, as three of the five weed populations we genotyped also have the highest expected heterozygosites that we measured, and the weedy populations clustered together in our genetic analyses. Empirical work also suggests that newly escaped radish cultivars would make poor weeds; Campbell and Snow (2009) found no evidence that *R. sativus* could establish feral populations without introgression from weedy radish and were unable to artificially select for greatly reduced flowering time in the Red Silk cultivar. While there are some reports of feral crop radish in the literature, it seems likely that these are actually hybrids (Snow and Campbell, 2005). Taken together, these results are inconsistent with an "escaped crop" or "crop hybrid" origin for weedy radish.

Our STRUCTURE and SmartPCA results suggest complex relationships between weedy *R.r. raphanistrum* and various populations of native *R.r. raphanistrum*. SmartPCA shows that the weeds are somewhat more similar to Eastern native populations, while the STRUCTURE results show almost no shared ancestry between those groups, but that two of the agricultural weed samples (COAU and WEAU) are more closely related to the Western native range *R.r. raphanistrum* (Figs. 1-1 and 1-2). Clustering by pairwise Fst suggests that the weeds are equally genetically distant from Eastern and Western natives, but that the natives are a single group (Fig. 1-3). Taken together, these results support the hypothesis that weedy radish is most closely related to native *R.r. raphanistrum*; however, it is impossible to say whether native Eastern or Western populations are more closely related to the weeds. Resolving these fine-grain relationships will likely require much more genetic and ecological data, as seeds can be dispersed long distances by contaminated grain shipments, seed predators, and livestock (Cheam,

2006; Holm, 1997; Snow and Campbell, 2005), and likely for this reason, *Raphanus* weed populations lack isolation by distance (Barnaud et al., 2013a,b; Kercher and Conner, 1996).

Our data are most consistent with native *R.r. raphanistrum* as the ancestor of weedy radish. A previous study found that weedy *R.r. raphanistrum* is indistinguishable from *R.r. landra* (previously *R.r. maritimus*) at mitochondrial markers (Yamagishi and Terachi, 2003), but the weeds have almost no overlap with *R.r. landra* populations in our SmartPCA analysis. Similarly, STRUCTURE shows only a few weedy individuals with minor introgression from *R.r. landra*, making the evolution of weedy *R.r. raphanistrum* directly from *R.r. landra* unlikely. Instead, both STRUCTURE and SmartPCA show some shared genetic background between weedy and native *R.r. raphanistrum* populations. Although our AMOVA found that weedy and native *R.r. raphanistrum* are genetically distinct, only 10% of the variation was accounted for by sub-species, compared to 13% and 19.5% for *R.r. sativus* and *R.r. landra* respectively (Fig. 1-3). These data are also in agreement with a previous phylogenetic analysis based on eight transcriptomes, which found that weedy *R.r. raphanistrum* and native *R.r. raphanistrum* are sister taxa, and that both are more closely related to the other native *Raphanus* than to any cultivar (Shen et al., 2013).

While these results are consistent with native *R.r. raphanistrum* as the ancestor of the weeds, we cannot rule out introgression from other *Raphanus* taxa, as the non-native populations are in the middle of all these taxa in PCA space, and our weed populations are highly heterozygous. Hybridization seems to be an important catalyst for invasive species in general (Ellstrand and Schierenbeck, 2000; Rius and Darling, 2014; Whitney and Gering, 2015); and *Raphanus* species are highly interfertile (Bett and Lydiate, 2003), readily interbreed, and there are multiple reports of hybridization in nature (Campbell et al., 2006; Hegde et al., 2006; Panetsos and Baker, 1967;

Snow and Campbell, 2005). However, all published reports of hybridization in *Raphanus* are between crops and populations of *R.r. raphanistrum* that are already weedy. While these instances cannot be used to determine the origin of the weed, they do illustrate the general importance of hybridization in this system.

Perhaps not surprisingly, *R. pugioniformis* (Ziffer-Berger et al., 2014) does not cluster genetically with any of the *R. raphanistrum* sub-species. In all runs where K is at least 3, STRUCTURE classifies *R. pugioniformis* as a unique group, and it falls into an intermediate region between native *R.r. raphanistrum* and the crops in genetic space (Fig. 1-1). However, it is similar to native *R.r. raphanistrum* and the seed crops in terms of average number of days to flower and vernalization requirement (Figs. 1-4 and A1-5). Perhaps most strikingly, *R. pugioniformis* is the only wild species to have anthocyanin floral pigments, a trait that otherwise characterizes the crops (unpub. data). This introduces the possibility that *R. sativus* was originally domesticated from *R. pugioniformis*, at least in part, although our data cannot distinguish between this scenario and one where *R. pugioniformis* is a crop-native hybrid, or admixed with an unknown population (Falush et al., 2016). A much larger genetic study, with increased marker density and more native populations, may help resolve these relationships.

Although STRUCTURE and SmartPCA gave similar answers, combining the results also led to some unexpected conclusions for crop origins. Superficially, our STRUCTURE plot supports the finding of many authors that crop radish was domesticated multiple times (Kong et al., 2011; Lü et al., 2008; Muminović et al., 2005; Wang et al., 2008; Yamagishi and Terachi, 2003; Yamane et al., 2005, 2009), as three of the four cultivars types have genetic backgrounds distinct from any other sampled *Raphanus*. (Fig. 1-2). However, whereas SmartPCA maintains some measure of genetic distance, STRUCTURE shows only statistically significant groupings. This means that

although the crop groupings in STRUCTURE are equally distinguishable, they are not necessarily similarly divergent from each other. Cultivars in our SmartPCA analysis, by contrast, form a small group in the space defined by the first two eigenvectors with the daikon cultivars clustered toward one edge (Fig. 1-1), which is more consistent with the single domestication event suggested by the transcriptome phylogeny in Shen et al. (2013). This pattern is also similar to that reported from a recent marker-based analysis of *R. sativus* wherein the crops also occupied a single swath of PCoA space, although those authors suggested their data was consistent with multiple origins (Kang et al., 2016).

This difference is likely largely due to sampling. Whereas Shen et al. (2013) and the current study assayed individuals across the genus and genome, previous studies either included only crop cultivars in their analysis (Kong et al., 2011; Muminović et al., 2005; Wang et al., 2008), or based their differentiation exclusively on chloroplast or mitochondrial DNA (Lü et al., 2008; Yamagishi and Terachi, 2003; Yamane et al., 2005, 2009) and so have sampled only a fraction of the genetic variation of *Raphanus*. In our SmartPCA results, the differentiation within the cultivars is dwarfed by the variation between crops and the rest of *Raphanus* (Figs. 1-1 and 1-2). This finding is consistent with single domestication event for *R. sativus*, followed by later radiation, but an explicitly phylogenetic analysis at a large number of markers is necessary to test this hypothesis.

**Evolution of faster flowering in weedy radish**

Although non-native *R.r. raphanistrum* populations are genetically similar to native *R.r. raphanistrum*, they flower much faster (25-58 days earlier flowering on average, Model 1). Weedy radish flowers much more rapidly and uniformly than any of the other *Raphanus* taxa,

and the weeds never require vernalization (Fig. 1-4). This lack of variation and decreased mean suggests that the weeds have undergone strong directional selection for flowering time. This supports the hypothesis that the weedy radish most likely arose from native *R.r. raphanistrum*, and subsequently evolved a faster flowering time.

This difference is especially striking in the raw data. We assayed days to flowering for non-native radish from three continents in common garden experiments across a span of eleven years and eight experiments across three locations, however we find almost no variation in the raw flowering time among or within the weed populations (Figs. 1-4, A1-8). This finding is somewhat surprising as phenotypic plasticity is expected to be a common feature of weeds (Baker, 1965) and invasives (Davidson et al., 2011; Richards et al., 2006) and flowering time has been found to be plastic in invasive plants e.g. (e.g. Claridge and Franklin, 2002; Colautti and Barrett, 2010). As our common garden experiments were performed either in the summer or using summer conditions, it is possible that they were simply not different enough to trigger a plastic response. However, some researchers have found directional selection for flowering time among invasives (Blair and Wolfe, 2004; Thurber et al., 2014), and weedy radish can be artificially selected for faster flowering times (Ashworth et al. 2016).

In stark contrast to the phenotypic uniformity of the weeds, native range populations of both *R.r. landra* and *R.r. raphanistrum* varied in both days to flowering, and the need for vernalization. Interestingly, this also appears not to be a plastic response to our variable conditions. In native populations that were assayed in at least four common garden experiments, we find far more variation within experiments than between them, and that some native populations have reproducibly bimodal distributions of flowering times (Fig. A1-8). Taken together, these data suggest that the native ancestors were likely more variable and slower-flowering, at least partly

due to a vernalization requirement, and this is consistent with the hypothesis of recent and rapid directional selection for shortened weed flowering time.

One possible mechanism for this rapid evolutionary change is introgression from other *Raphanus* taxa, and hybridization has preceded changes flowering time in *Raphanus* in several cases (Campbell and Snow, 2007, 2009; Campbell et al., 2009a; Hegde et al., 2006; Snow et al., 2010). Similarly, both natural and artificial crop-weed radish hybrids have been found to have increased competitive ability, at least under some conditions (Campbell and Snow, 2007; Campbell et al., 2006; Hegde et al., 2006). Most notably, California wild radish is a hybrid of *R.r. raphanistrum* and various cultivars which has invaded large swaths of central California (Panetsos and Baker, 1967). Only a hundred years after its hybridization, California wild radish has extirpated the local weedy radish, and shows differentiation and local adaptation in several life history traits, including changes in flowering time (Hegde et al., 2006; Ridley and Ellstrand, 2010), which suggests that hybridization can induce weed invasion.

However, all these examples of hybridization in *Raphanus* are hybrids between crops and established weeds, and so cannot have contributed to the creation or phenotypic divergence of weedy radish. Still, some comparisons can be made, and our results suggest that large-scale hybridization is probably not important in weedy radish evolution. In naturally occurring hybrids such as California wild radish, STRUCTURE analysis showed a clear pattern of admixture in each hybrid individual (Hegde et al., 2006), whereas our STRUCTURE results assign most individual weeds to a single genetic background, even at very high K (data not shown). Furthermore, in most of these studies, hybridization resulted in more variable (Hegde et al., 2006) or later flowering times, (Campbell and Snow, 2007; Campbell et al., 2009a; Snow et al., 2010, but see: Campbell and Snow, 2009) which is not what we find in our weed populations.

However, Snow et al. (2010) found that some crop alleles persisted for several years after experimental hybridization, albeit at lower frequencies than expected. As the marker density of our data is relatively low, we would be unable to resolve small scale introgression followed by strong selection for adaptive alleles. Denser markers across a similarly broad collection of populations would be useful both for detecting potential introgression events and verifying that the weeds evolved from a single native source population.

In summary, we found no evidence to support a crop origin for weedy radish, either by hybridization or exoferality. Non-native *R.r. raphanistrum* likely descended from native *R.r. raphanistrum*, with possible introgression from other *Raphanus* taxa. All of these potential source populations have longer flowering times than we found among the weeds, which suggests that wild populations were not pre-adapted to field conditions and is evidence for rapid local adaptation to an agricultural habitat. These results were not unexpected, as rapid life cycle is a common feature of weeds (Baker, 1965), however the lack of phenotypic variation in our weed populations is both striking and unexpected. Weedy radish has been artificially selected to flower both substantially faster and slower in only five generations (Ashworth et al., 2016), and is both highly heterozygous and an obligate outcrosser, which suggests that there is not a corresponding lack of genetic variation for flowering time loci. Our study did not explicitly test for plasticity, however the lack of variation among weed populations across our common garden locations suggests that the weeds are not plastic for flowering time.

This is unexpected because plasticity is much more common in weedy or invasive plants than in their native counterparts (Davidson et al., 2011), and plasticity of reproductive traits is a frequent feature of weedy and invasive plants (e.g., Atlan et al., 2015; Chaney and Baucom, 2012; Meerts, 1995; Novy et al., 2012), and some other members of Brassicaceae (Franks et al., 2007).

However, while observations of common weeds and invasive plants provide a good framework for agricultural weed research, they are not directly comparable. Weeds and invasives are expected to favor plasticity over local adaptation because they are highly mobile, and so inhabit a wide range of habitats (Palacio-López et al., 2015). Agricultural weeds are frequently moved around the globe via contaminated seed stock, however the habitats they move to are relatively uniform. We might therefore expect plasticity in flowering time to be maladaptive in a highly regimented ecosystem such as an active farm. Whether or not directional selection is a more general phenomenon in agricultural conditions requires additional studies comparing key traits between agricultural weeds and their progenitors.

**Data Accessibility**

Seed stock information is in Table S1. Genotypic data is available at Dryad (doi: TBD). Scripts and data required for all analyses are available via the authors github page.

**APPENDIX**

| G | P | Population | Variety/Range | Habitat | Source Site | Stock/Collection location |
|---|---|---|---|---|---|---|
| **_R.r.landra_** | | | | | | |
| X | X | PBFR | Native | Undisturbed | France | 42°48'690N, 3°02'010E |
| X | X | CBES | Native | Undisturbed | Spain | 43.4°N 4.23°W |
| X | X | SAES | Native | Undisturbed | Spain | 43.49°N, 3.53°W |
| **_R.r.raphanistrum_** | | | | | | |
| X | X | AFFR | Western | Ag field | France | 43°08'260"N, 2°53'547"E |
| X | X | DEES | Western | Undisturbed | Spain | 38°23'25.18"N, 3°29'39.88"W |
| X | X | HCES | Western | Disturbed | Spain | 37°18'05.93"N, 5°57'57.81"W |
| X | X | HMES | Western | Disturbed | Spain | 37°16'35.66"N, 5°57'16.80"W |
| X | X | IMES | Western | Disturbed | Spain | 37°13'16.73"N, 5°58'30.50"W |
| X | X | MAES | Western | Disturbed | Spain | 43°29'31.859"N, 3°31'28.808"W |
| X | X | GHIL | Eastern | Disturbed | Israel | 32°10'30.5"N, 34°56'01.6"E |
| X | X | HZIL | Eastern | Disturbed | Israel | 32°10'30.5"N, 34°49'28.9"E |
| X | X | REIL | Eastern | Disturbed | Israel | 31°54.157'N, 34°49.407'E |
| X | X | TYIL | Eastern | Disturbed | Israel | 32°15.284'N, 34°52.015'E |
| X | X | ZYIL | Eastern | Disturbed | Israel | 31°56'00.4"N, 34°47'11.4"E |
| X | X | AUFI | Non-Native | Ag field | Finland | 60°38'54"N, 22°33'53"E |
| X | X | BINY | Non-Native | Ag field | New York | 42°11'2.4"N, 75°50'7.08"W |
| X | X | COAU | Non-Native | Ag field | Australia | 31°18'S, 152°20'E |
| | X | KAMI | Non-Native | Ag field | Michigan, USA | 42°21'34.6"N, 85°35'54.2"W |
| | X | MAFI | Non-Native | Disturbed | Finland | 60°33'12"N, 22°06'53"E |
| | X | NAAU | Non-Native | Ag field | Austraila | 36°96'S, 140°73'E |
| X | | NCDE | Non-Native | Disturbed | Germany | 52°58.698'N, 9°37.865'E; alt: 224 |
| X | X | WEAU | Non-Native | Ag field | Austraila | 31°23'S, 118°32'E |
| **_R.pugioniformis_** | | | | | | |
| X | X | GMIL | Native Range | Undisturbed | Israel | 32°30'01.6"N, 35°24'51.4"E |
| **_R.s.convar.sativus_** | | | | | | |
| | X | CGBC | Chinese Green Luobo | daikon | Baker Creek Heirloom | Cat#RD119 |
| | X | FGBC | Formosa Giant Luobuo | daikon | Baker Creek Heirloom | Cat#RD127 |
| X | X | MYJO | Miyashige | daikon | John Scheepers | 625.11 |
| X | X | NEJS | New Crown | daikon | John Scheepers | 3860 |
| X | X | TOBG | All Seasons Tokinashi | daikon | Bountiful Gardens | VRA-5050 |
| X | X | WMBG | Watermelon | daikon | Bountiful Gardens | VRA-5100 |
| X | X | CBBG | Cherry Belle | European | Bountiful Gardens | VRA-5080 |
| X | X | DAJO | D'avignon | European | John Scheepers | 620.11 |
| X | X | ESNK | Early Scarlet Globe | European | NK Lawn & Garden Co | 7576 |
| | X | FRSI | Flamboyant Long Italian | European | GrowItalian.com | |
| | X | LBBC | Long Black Spanish | European | Baker Creek Heirloom | |
| | X | RABS | Raxe | European | Burpee's Signature Seeds | |
| | X | RBBC | Round Black Spanish | European | Baker Creek Heirloom | |
| X | X | SPNK | Sparkler | European | NK Lawn & Garden Co | 7583 |
| **_R.s.convar.oleifera_** | | | | | | |
| X | X | ADOL | Adagio | Oilseed | MSU | |
| X | X | AROL | Arena | Oilseed | MSU | |
| X | X | COOL | Colonel | Oilseed | MSU | |
| | X | OIBG | Oilseed Radish | Oilseed | Bountiful Gardens | GRA-7378 |
| **_R.s.convar.caudatus_** | | | | | | |
| X | X | MABG | Madras | Rat tail | Bountiful Gardens | VRA-5060 |
| X | X | RABG | Rat's Tail | Rat tail | Bountiful Gardens | VRA-5070 |
| X | X | RAJS | Rat's Tail | CRat tail | John Scheepers | 3870 |

Figure A1-1: All populations were included in at least one analysis of the paper: Genetic(G) or Phenotypic(P). Variety/Range refers to whether a wild population was collected inside (Native Range) or outside the Mediterranean region; or to the convariety name for cultivated varieties. Source refers to the individual who collected the source populations in the case of wild populations, or the company that cultivar seed was purchased from. Stock/Collection location is the seed company stock number for cultivars, and global position of the source population in the case of wild collected plants.

| Locus | Linkage | Cm | Contig | Enzyme | Forward Primer | Reverse Primer | Amp/Rep | Allele 1 | Allele 2 | Mg(mM) | Tm(C) | Anneal(C) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DWRD_124 | 1 | 0 | CL427Contig8 | HindIII | TGGCGGAAAGCAAAGAGAACTACG | AAAGGAAAGTCACAAGCGGTGCAG | 400 | 250 | 150 | 2.0 | 58 | 72 |
| Na12-E05 | 1 | 12 | | | CGTATGTTTGTTCCACCTGC | ACTAGCAACCACAACGGACC | CA | 133 | 150 | 2.5 | 55 | 52 |
| DWRD_112 | 1 | 37 | CL22Contig17 | HindIII | TGACCTTGACCTTGATTCCGAGCA | ATGTTCTCGGTGAGAAAGGGAGGA | 1200 | 1000 | 200 | 2.0 | 58 | 72 |
| Na14-E08 | 1 | 48 | | | TTACTATCCCCTCTCCGCAC | GCGGATTATGATGACGCAG | GCC | 85 | 122 | 2.5 | 55 | 50 |
| DWRD_61 | 1 | 54 | CL2272Contig2 | PstI | TAGTGGTTCGTCATCGGCTTCAGT | TAACTCACTTGTTGCCGGAGCAGA | 700 | 600 | 100 | 2.0 | 58 | 72 |
| DWRD_123 | 2 | 2 | CL4189Contig2 | HindIII | CCTTTGAGCTTGCGCTTTCCTTCT | TACCCACTTGGATGGCAGAAACCT | 400 | 225 | 175 | 2.0 | 58 | 72 |
| DWRD_107 | 2 | 16 | CL1985Contig1 | HindIII | AAACCGGTTCCATGAGAATGCCAC | ATCCTGGGCAGCTCAATTACCCAA | 370 | 200 | 170 | 2.0 | 58 | 72 |
| DWRD_177 | 2 | 46 | CL2355Contig3 | BcII | ATCATTCTCATCCTCAGTCGCCCT | AGTTTGTGAGGACGAGGCTTGACT | 600 | 400 | 200 | 2.0 | 58 | 72 |
| Ra1-H08 | 3 | 9 | | | GTCGATGATCACGGAAGAGG | CTTGACAGCTACGGTTTGTCC | AGG/CGG | 184 | 204 | 2.5 | 56 | 52 |
| Bn26A | 5 | 42 | | | TAAACTTGTCAGACGCCGTTATC | CCCGTAAATCAAGCAAATGG | GA | 90 | 120 | 2.5 | 53 | 52 |
| BRMS-005 | 5 | 57 | | | ACCTCCTGCAGATTCGTGTC | GCTGACCTTTCTTACCGCTC | GA | 132 | 175 | 2.5 | 56 | 50 |
| DWRD_121 | 6 | 11 | CL370Contig1 | HindIII | TCATCTTCCTCCTCGGTTGCTGAT | AGAAATCGACCGGATGGTGAAGGA | 400 | 250 | 150 | 2.0 | 58 | 72 |
| DWRD_48 | 6 | 23 | CL1174Contig1 | PstI | CACCACCGCCCAATCTCAACAAAT | TGGTATAGCAAGGGCAGCGTAAGT | 600 | 400 | 200 | 2.0 | 58 | 72 |
| Ra2-E11 | 6 | 53 | | | GGAGCCAGGAGAGAAGAAGG | CCCAAAACTTCCAAGAAAAGC | CT | 167 | 191 | 2.5 | 54 | 61 |
| DWRD_158 | 6 | 65 | CL3595Contig2 | EcoRi | CAAGCCGCAGACCAAATCAACACT | GCTGCCGAGCTTGAACCAAACATT | 400 | 250 | 150 | 2.0 | 58 | 72 |
| DWRD_180 | 8 | 21 | CL2638Contig1 | BcII | ATCCAACTTGACGGTGTCAACGGA | AAAGGTGTTCGTGTCTGGCTAGGT | 350 | 250 | 100 | 2.0 | 58 | 72 |
| DWRD_205 | 9 | 20 | CL1128Contig1 | EcoRV | GTGGTTTCGAAGCTTTGTTTCTCCG | TAGTTGTCGGGAGGAAGCGTGATT | 300 | 275 | 25 | 2.0 | 58 | 72 |
| DWRD_27 | 9 | 35 | CL126Contig10 | NsiI | AGGTCCGGCTTCTCTAGTGATCTT | ACTCTGTCAAGTCATGCTTCGCGT | 600 | 400 | 200 | 2.0 | 58 | 72 |
| Na10-H06 | S1 | | | | AGAATGAGACCCAGAAACCG | GCCACACTCTCTCTTACTAGGGC | CT | 120 | 140 | 2.5 | 56 | 52 |
| DWRD_97 | S2 | | CL1509Contig2 | HindIII | TGACGTGTAGTGTAGCGTTTGCGT | TGAACATAGAACCGACCACCTCCA | 650 | 350 | 300 | 2.0 | 58 | 72 |
| Bn35d | S3 | | | | GCAGAAGGAGGAGAAGAGTTGG | TTGAGCCGTAAAGTTGTCACCT | GA | 222 | 251 | 2.5 | 58 | 61 |

Figure A1-2: PCR parameters and restriction enzymes for 21 SSR and CAPS markers. The allele 1 and 2 and are the sizes of the two fragments for the CAPS markers, and the size of the largest and smallest possible alleles for the SSRs.
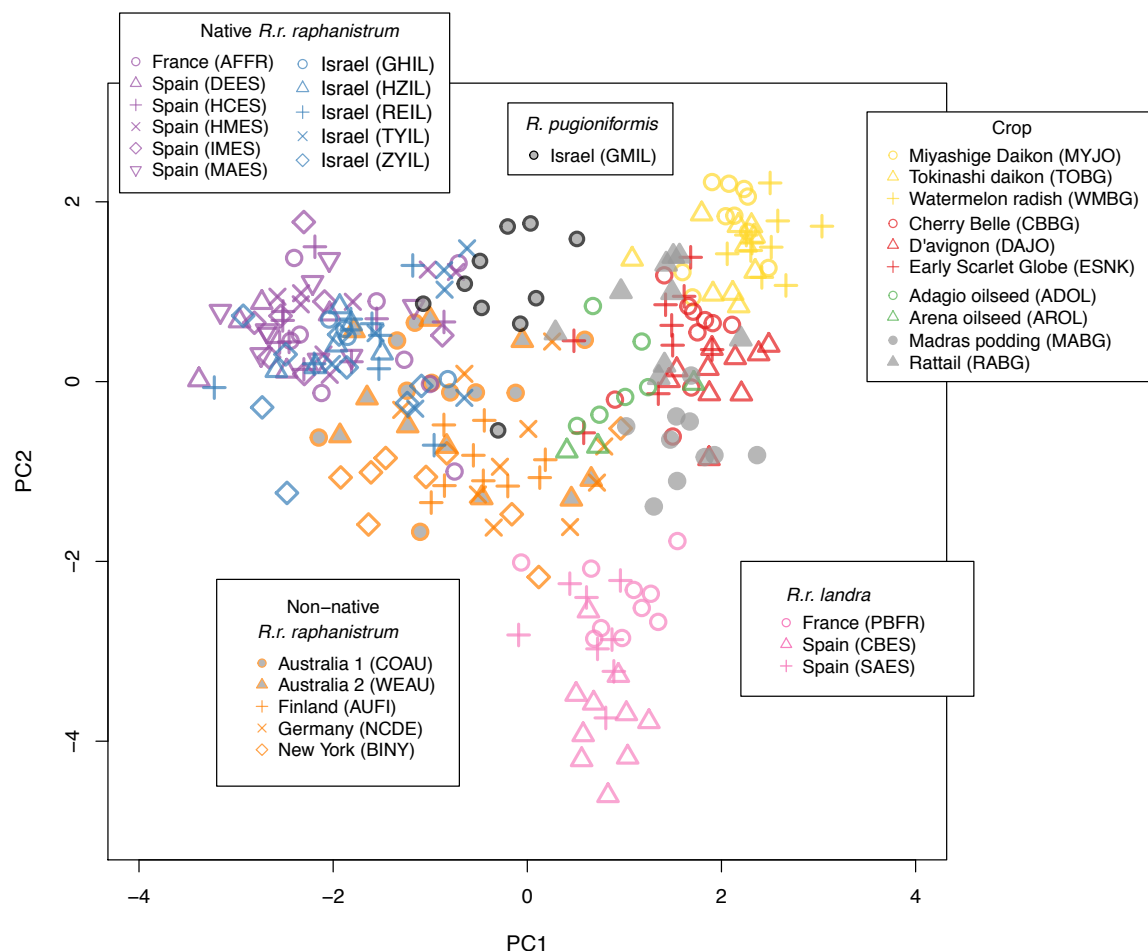
Figure A1-3: PCA plot of the first two eigenvectors of a principal components analysis of 30 *Raphanus* populations genotyped at presumed neutral markers. Each point is an individual, and each population is represented by 8-10 individuals. Populations are identified by plotting character and colored to match the SmartPCA figure in the main text (Fig. 1-1) This PCA was generated in R by running prcomp on the same dataset given to SmartPCA but with missing data omitted (na.omit). This PCA largely recapitulates the SmartPCA results, however some individuals have been lost due to missing data
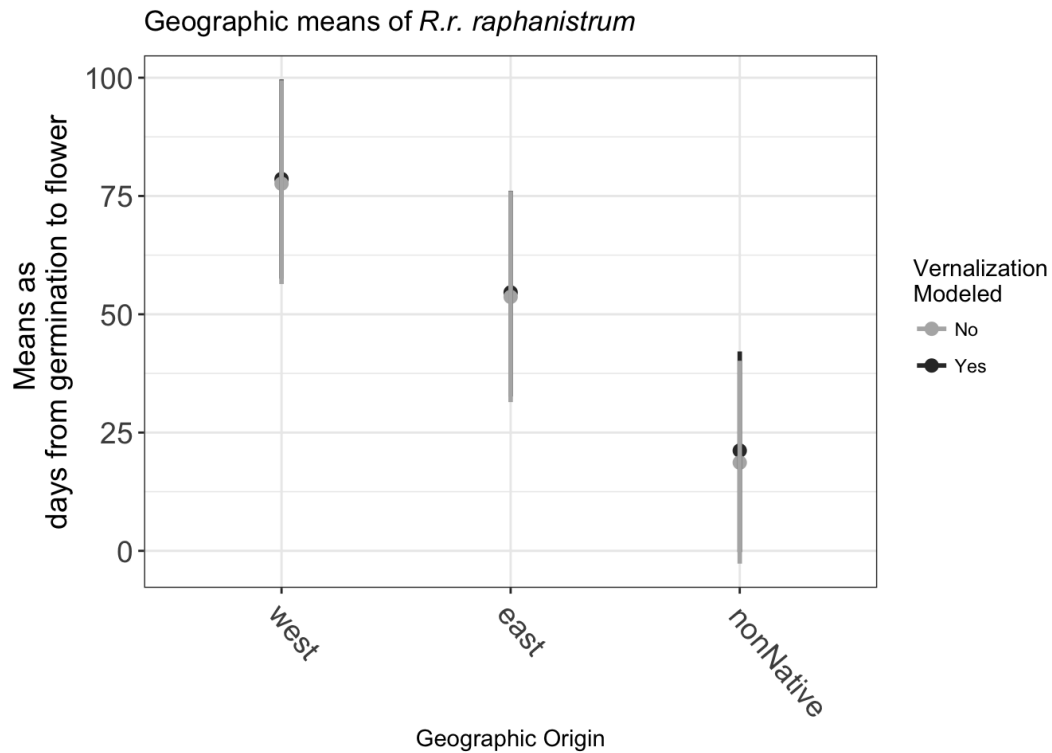
Figure A1-4: Geographic means of *R.r. raphanistrum* and 95% credible intervals for number of days from germination to flowering. 'NonNative' populations were collected outside of the Mediterranean region, all but one from agricultural fields (Fig. A1-1). 'East' and 'West' populations were collected either in Israel or in Spain or France, respectively. Removing the fixed effect of vernalization from the model had a negligible effect on geographic means.
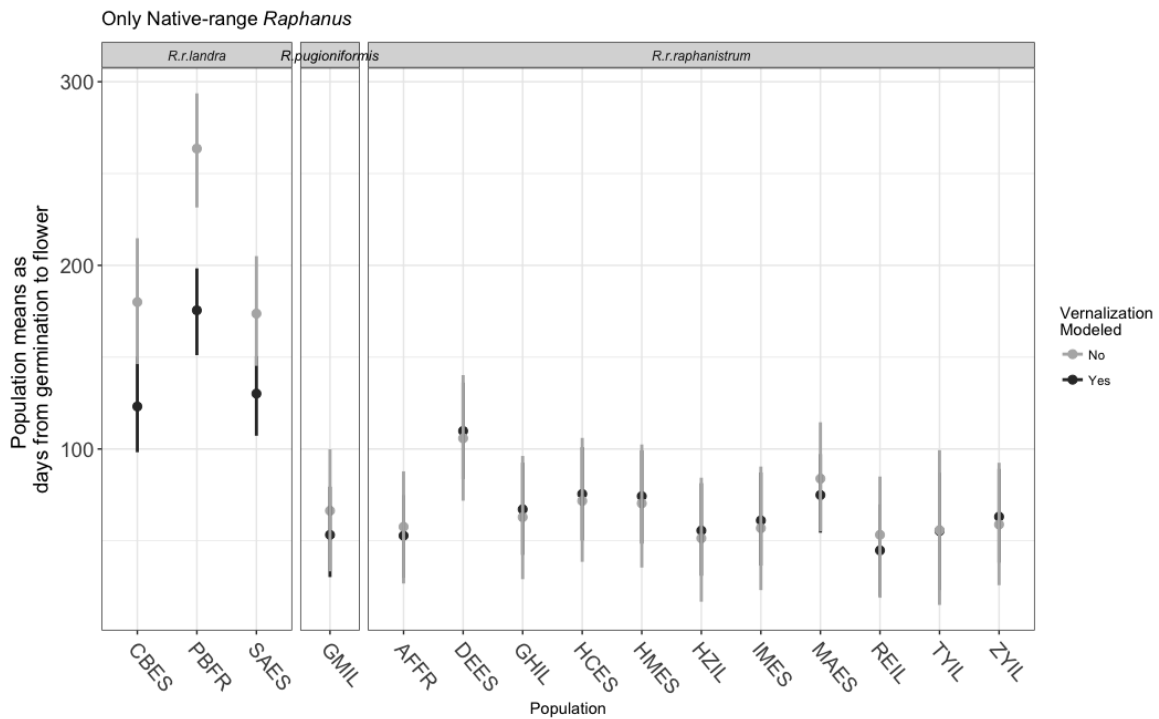
Figure A1-5: Native range *Raphanus* population means and 95% credible intervals for number of days from germination to flowering for all *Raphanus* populations collected in the Mediterranean region. Accounting for differences in vernalization treatment had a negligible effect in most instances, however it significantly reduced the estimates for *R.r. landra* populations. (Population abbreviations as per Fig. A1-1)
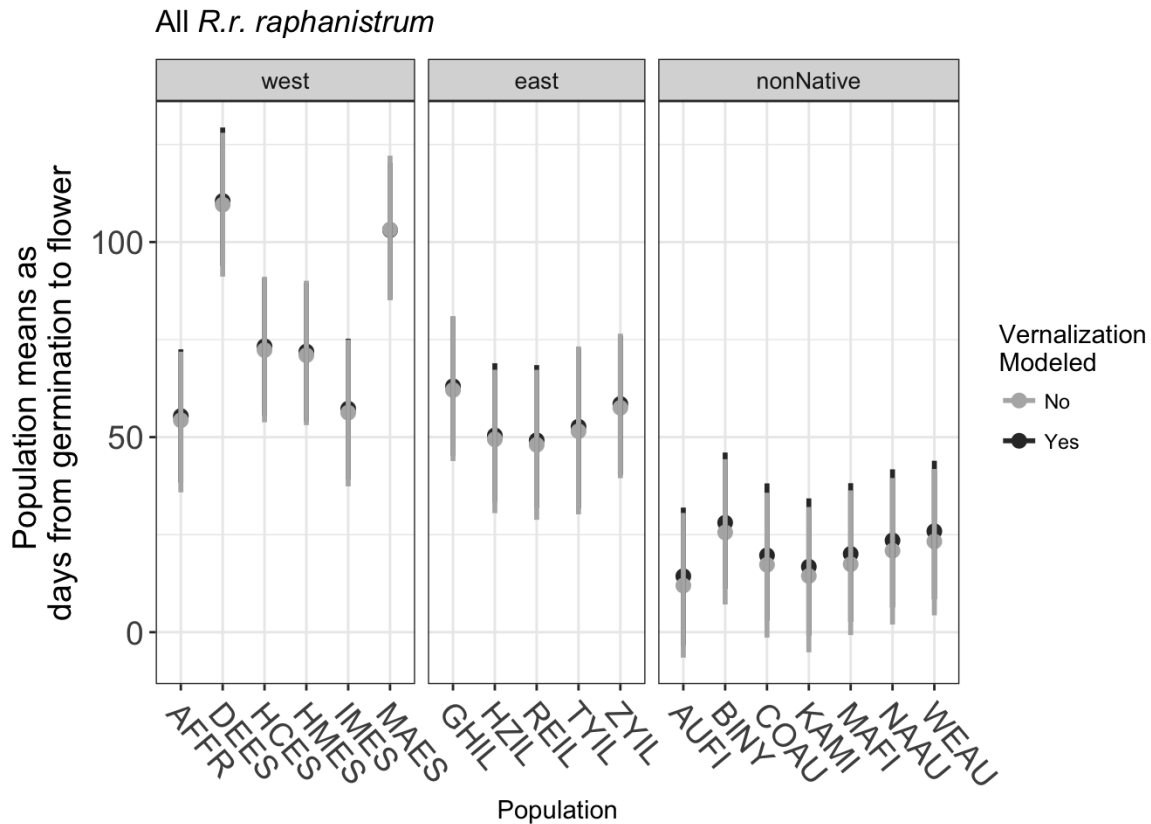
Figure A1-6: Population means of *R.r. raphanistrum* and 95% credible intervals for no. days from germination to flowering. 'NonNative' populations were collected from agricultural fields outside of the Mediterranean region. 'East' and 'West' populations were collected either in Israel or in Spain or France, respectively. Removing the fixed effect of vernalization from the model had a negligible effect on population means.
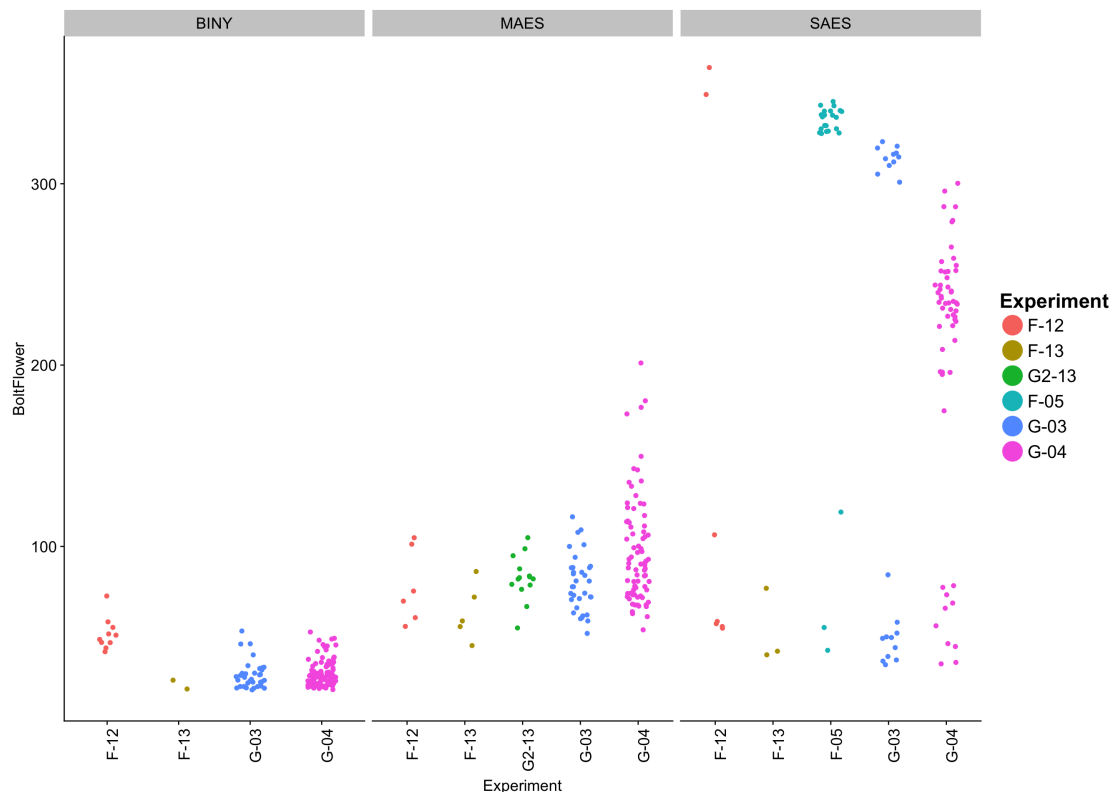
Figure A1-8: Raw flowering times of all individuals from three populations. BINY, a weed, shows little variation in flowering time within or between experiments. Both MAES and SAES show a large amount of within population variation in Figure 1-4, however very little of this variation appears to be due to experiment. SAES has very similar flowering time distributions in all experiments except F-13, where it appears they all flowered early. However, F-13 was a field experiment, and none of the SAES plants survived the winter, so individuals who were alive and would have likely flowered the following year are not included. Had they survived, SAES would have had a bi-modal flowering time distribution in all experiments

| Population | Het | CoV |
|---|---|---|
| PBFR | 0.34 | 2.50 |
| CBES | 0.21 | 56.9 |
| SAES | 0.28 | 50.7 |
| *R.r.landra* | **0.35** | **44.0** |
| AFFR | 0.41 | 45.0 |
| DEES | 0.36 | 17.1 |
| HCES | 0.43 | 18.8 |
| HMES | 0.45 | 19.1 |
| IMES | 0.43 | 28.2 |
| MAES | 0.41 | 29.3 |
| *WesternR.r.raphanistrum* | **0.47** | **43.4** |
| GHIL | 0.38 | 17.1 |
| HZIL | 0.42 | 28.1 |
| REIL | 0.42 | 28.9 |
| TYIL | 0.44 | 31.6 |
| ZYIL | 0.39 | 20.3 |
| *EasternR.r.raphanistrum* | **0.45** | **27.8** |
| AUFI | 0.42 | 12.7 |
| BINY | 0.50 | 28.8 |
| COAU | 0.38 | 29.6 |
| KAMI | | 18.1 |
| MAFI | | 13.9 |
| NAAU | | 23.1 |
| NCDE | 0.47 | |
| WEAU | 0.48 | 24.1 |
| *WeedyR.r.raphanistrum* | **0.52** | **25.7** |
| GMIL | 0.36 | 24.8 |
| *R.pugioniformis* | **0.36** | **24.8** |
| CGBC | | 27.7 |
| FGBC | | 27.0 |
| MYJO | 0.33 | 37.4 |
| NEJS | 0.32 | 15.5 |
| TOBG | 0.26 | 18.2 |
| WMBG | 0.22 | 26.8 |
| *Daikon, R.s.convar.sativus* | **0.36** | **42.0** |
| CBBG | 0.28 | 24.6 |
| DAJO | 0.23 | 14.7 |
| ESNK | 0.26 | 26.7 |
| FRSI | | 18.4 |
| LBBC | | 24.8 |
| RABS | | 26.3 |
| RBBC | | 29.1 |
| SPNK | 0.24 | 36.8 |
| *European, R.s.convar.sativus* | **0.33** | **32.0** |
| ADOL | 0.34 | 31.9 |
| AROL | 0.39 | 19.2 |
| COOL | 0.44 | 21.3 |
| OIBG | | 21.8 |
| *R.s.convar.oleifera* | **0.45** | **31.7** |
| MABG | 0.38 | 18.0 |
| RABG | 0.35 | 20.2 |
| RAJS | 0.38 | 30.1 |
| *R.s.convar.caudatus* | **0.44** | **24.3** |

Figure A1-9: Selected summary statistics. Bolded lines are *apriori* grouping values and are preceded by population values for that group. 'Het' is the expected heterozygosity as calculated from all 21 markers. CoV is the coefficient of variation for days for flower, calculated as mean/standard deviation.

**REFERENCES**

# REFERENCES

Agrawal, A. A., Conner, J. K., Johnson, M. T. J., and Wallsgrove, R. (2002). Ecological genetics of an induced plant defense against herbivores: additive genetic variance and costs of phenotypic plasticity. *Evolution*, 56(11):2206–2213.

Ashworth, M. B., Walsh, M. J., Flower, K. C., Vila-Aiub, M. M., and Powles, S. B. (2016). Directional selection for flowering time leads to adaptive evolution in *Raphanus raphanistrum* (Wild radish). *Evolutionary Applications*, 9(4):619–629.

Atlan, A., Hornoy, B., Delerue, F., Gonzalez, M., Pierre, J.-S., and Tarayre, M. (2015). Phenotypic Plasticity in Reproductive Traits of the Perennial Shrub *Ulex europaeus* in Response to Shading: A Multi-Year Monitoring of Cultivated Clones. *PLoS One*, 10(9):1–17.

Baker, H. G. (1965). Characteristics and modes of origin of weeds. In Baker, H. G. and Stebbins, G. L., editors, *The genetics of colonizing species*, pages 147–172. Academic Press, New York & London.

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

Baker, H. G. (1991). The continuing evolution of weeds. *Economic Botany*, 45(4):445–449.

Barnaud, A., Kalwij, J. M., Berthouly-Salazar, C., McGeoch, M. A., and Jansen van Vuuren, B. (2013a). Are road verges corridors for weed invasion? Insights from the fine-scale spatial genetic structure of *Raphanus raphanistrum*. *Weed Research*, 53(5):362– 369.

Barnaud, A., Kalwij, J. M., McGeoch, M. A., and van Vuuren, B. J. (2013b). Patterns of weed invasion: evidence from the spatial genetic structure of *Raphanus raphanistrum*. *Biological Invasions*, 15(11):2455–2465.

Barrett, S. C. H. (1983). Crop mimicry in weeds. *Economic Botany*, 37(3):255–282.

Bett, K. E. and Lydiate, D. J. (2003). Genetic analysis and genome mapping in *Raphanus*. *Genome*, 46(3):423–430.

Blackshaw, R. E., Lemerle, D., Mailer, R., and Young, K. R. (2002). Influence of wild radish on yield and quality of canola. *Weed Science*, 50(3):344–349.

Blair, A. C. and Wolfe, L. M. (2004). The evolution of an invasive plant: An experimental study with *Silene latifolia*. *Ecology*, 85(11):3035–3042.

Buswell, J. M., Moles, A. T., and Hartley, S. (2011). Is rapid evolution common in introduced plant species? *Journal of Ecology*, 99(1):214–224.

Campbell, L. G. and Snow, A. A. (2007). Competition alters life history and increases the relative fecundity of crop-wild radish hybrids ( *Raphanus* spp.). *New Phytologist,* 173(3):648–660.

Campbell, L. G. and Snow, A. A. (2009). Can feral weeds evolve from cultivated radish (*Raphanus sativus*, Brassicaceae)? *American Journal of Botany*, 96(2):498–506.

Campbell, L. G., Snow, A. A., and Ridley, C. E. (2006). Weed evolution after crop gene introgression: greater survival and fecundity of hybrids in a new environment. *Ecology Letters*, 9(11):1198–1209.

Campbell, L. G., Snow, A. A., and Sweeney, P. M. (2009a). When divergent life histories hybridize: insights into adaptive life-history traits in an annual weed. *New Phytologist,* 184(4):806–818.

Campbell, L. G., Snow, A. A., Sweeney, P. M., and Ketner, J. M. (2009b). Rapid evolution in crop-weed hybrids under artificial selection for divergent life histories. *Evolutionary Applications,* 2(2):172–186.

Chaney, L. and Baucom, R. S. (2012). The evolutionary potential of Baker's weediness traits in the common morning glory, *Ipomoea purpurea* (Convolvulaceae). *American Journal of Botany*, 99(9):1524–1530.

Cheam, A. H. (2006). *Proceedings of the Wild Radish and Other Cruciferous Weeds Symposium*. Department of Agriculture and Food, Western Australia.

Claridge, K. and Franklin, S. B. (2002). Compensation and plasticity in an invasive plant species. *Biological Invasions*, 4(4):339–347.

Colautti, R. I. and Barrett, S. C. H. (2010). Natural Selection and Genetic Constraints on Flowering Phenology in an Invasive Plant. *International Journal of Plant Sciences*, 171(9):960–971.

Conner, J. K. (2002). Genetic mechanisms of floral trait correlations in a natural population. *Nature*, 420(6914):407–410.

Conner, J. K., Franks, R., and Stewart, C. (2003). Expression of additive genetic variances and covariances for wild radish floral traits: comparison between field and greenhouse environments. *Evolution*, 57(3):487–495.

Conner, J. K., Karoly, K., Stewart, C., Koelling, V. A., Sahli, H. F., and Shaw, F. H. (2011). Rapid Independent Trait Evolution despite a Strong Pleiotropic Genetic Correlation. *The American Naturalist*, 178(4):429–441.

Conner, J. K., Mills, C. J., Koelling, V. A., and Karoly, K. (2014). Artificial selection on anther exsertion in wild radish, *Raphanus raphanistrum*. *Scientific Data*, 1:140027.

Culpepper, S. (2012). Managing Wild Radish in Wheat. Technical report, University of Georgia.

Davidson, A. M., Jennions, M., and Nicotra, A. B. (2011). Do invasive species show higher phenotypic plasticity than native species and, if so, is it adaptive? A meta-analysis. *Ecology Letters*, 14(4):419–431.

de Wet, J. (1966). The Origin of Weediness in Plants. *Proc. of the Okla. Acad. of Sci.*, pages 14–17.

de Wet, J. and Harlan, J. R. (1975). Weeds and domesticates: evolution in the man-made habitat. *Economic Botany*, 29(2):99–107.

Dekker, J. (2011). *Evolutionary Ecology of Weeds*. Weed Biology Laboratory, 1.1.11 edition.

Desplanque, B., Boudry, P., Broomberg, K., Saumitou-Laprade, P., Cuguen, J., and Van Dijk, H. (1999). Genetic diversity and gene flow between wild, cultivated and weedy forms of *Beta vulgaris L*. (Chenopodiaceae), assessed by RFLP and microsatellite markers. *Theoretical and Applied Genetics*, 98(8):1194–1201.

Devlin, B. and Ellstrand, N. C. (1990). Male and female fertility variation in wild radish, a hermaphrodite. *The American Naturalist*, 136(1):87–107.

Earl, D. A. and vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2):359–361.

Ellstrand, N. C., Heredia, S. M., Leak-Garcia, J. A., Heraty, J. M., Burger, J. C., Yao, L., Nohzadeh-Malakshah, S., and Ridley, C. E. (2010). Crops gone wild: evolution of weeds and invasives from domesticated ancestors. *Evolutionary Applications*, 3(5-6):494–504.

Ellstrand, N. C., Meirmans, P., Rong, J., Bartsch, D., Ghosh, A., de Jong, T. J., Haccou, P., Lu, B.-R., Snow, A. A., Stewart Jr, C. N., Strasburg, J. L., Van Tienderen, P. H., Vrieling, K., and Hooftman, D. (2013). Introgression of Crop Alleles into Wild or Weedy Populations. *Annual Review of Ecology, Evolution, and Systematics*, 44(1):325–345.

Ellstrand, N. C. and Schierenbeck, K. A. (2000). Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences*, 97(13):7043–7050.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, 14(8):2611–2620.

Falush, D., van Dorp, L., and Lawson, D. (2016). A tutorial on how (not) to over-interpret STRUCTURE/ADMIXTURE bar plots. *bioRxiv*, pages 1–16.

Franks, S. J., Sim, S., and Weis, A. E. (2007). Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proceedings of the National Academy of Sciences*, 104(4):1278–1282.

Gressel, J. (2005). *Crop Ferality and Volunteerism*. CRC Press.

Guillot, G., Estoup, A., Mortier, F., and Cosson, J. F. (2005). A Spatial Statistical Model for Landscape Genetics. *Genetics*, 170(3):1261–1280.

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, 33(2):1–22.

Harlan, J. R. (1992). What Is a Weed? In *Crops & Man*, pages 83–99. American Society of Agronomy.

Hegde, S. G., Nason, J. D., Clegg, J. M., and Ellstrand, N. C. (2006). The evolution of California's wild radish has resulted in the extinction of its progenitors. *Evolution*, 60(6):1187–1197.

Higgins, R., Heikes, E., and Kempen, H. (1978). The Case for Preventitive Weed Man- agement. *Proceedings of the Western Society of Weed Science*, 31:35–40.

Holm, L. G. (1997). *World Weeds*. Natural Histories and Distribution. John Wiley & Sons Inc.

Irwin, R. E., Strauss, S. Y., Storz, S., Emerson, A., and Guibert, G. (2003). The role of herbivores in the maintenance of a flower color polymorphism in wild radish. *Ecology*, 84(7):1733–1743.

Kanapeckas, K. L., Vigueira, C. C., Ortiz, A., Gettler, K. A., Burgos, N. R., Fischer, A. J., and Lawton-Rauh, A. L. (2016). Escape to Ferality: The Endoferal Origin of Weedy Rice from Crop Rice through De-Domestication. *PLoS One,* 11(9):e0162676–23.

Kang, E. S., Ha, S. M., Ko, H. C., Yu, H.-J., and Chae, W. B. (2016). Reproductive traits and molecular evidence related to the global distribution of cultivated radish (*Raphanus sativus L.*). *Plant Systematics and Evolution*, 302(10):1367–1380.

Kercher, S. and Conner, J. K. (1996). Patterns of genetic variability within and among populations of wild radish, *Raphanus raphanistrum* (Brassicaceae). *American Journal of Botany,* 83(11):1416–1421.

Klinger, T., Elam, D., and Ellstrand, N. C. (1991). Radish as a Model System for the Study of Engineered Gene Escape Rates Via Crop-Weed Mating. *Conservation Biology*, 5(4):531–535.

Kong, Q., Li, X., Xiang, C., Wang, H., Song, J., and Zhi, H. (2011). Genetic Diversity of Radish (*Raphanus sativus L.)* Germplasm Resources Revealed by AFLP and RAPD Markers. *Plant Molecular Biology Reporter*, 29(1):217–223.

Lai, Z., Kane, N. C., Zou, Y., and Rieseberg, L. H. (2008). Natural Variation in Gene Expression Between Wild and Weedy Populations of *Helianthus annuus*. *Genetics*, 179(4):1881–1890.

Lehtilä, K. and Strauss, S. Y. (1999). Effects of Foliar Herbivory on Male and Female Reproductive Traits of Wild Radish, *Raphanus raphanistrum*. *Ecology*, 80(1):116–124.

Lü, N., Yamane, K., and Ohnishi, O. (2008). Genetic diversity of cultivated and wild radish and phylogenetic relationships among *Raphanus* and Brassica species revealed by the analysis of trnK/matK sequence. *Breeding Science*, 58(1):15–22.

Malik, M. S. (2009). *Biology and ecology of wild radish (Raphanus raphanistrum).* PhD thesis, ProQuest, Clemson University.

Mazer, S. J. and Schick, C. T. (1991). Constancy of population parameters for life-history and floral traits in *Raphanus sativus L.* II. Effects of planting density on phenotype and heritability estimates. *Evolution*, 45(8):1888–1907.

McNeill, J. (1976). The taxonomy and evolution of weeds. *Weed Research*, 16:399–413.

Meerts, P. (1995). Phenotypic Plasticity in the Annual Weed *Polygonum aviculare*. *Botanica Acta*, 108(5):414–424.

Moghe, G. D., Hufnagel, D. E., Tang, H., Xiao, Y., Dworkin, I., Town, C. D., Conner, J. K., and Shiu, S.-H. (2014). Consequences of Whole-Genome Triplication as Revealed by Comparative Genomic Analyses of the Wild Radish *Raphanus raphanistrum* and Three Other Brassicaceae Species. *The Plant Cell*, 26(5):1925–1937.

Morgan, M. T. and Conner, J. K. (2001). Using genetic markers to directly estimate male selection gradients. *Evolution*, 55(2):272–281.

Muller, M.-H., Latreille, M., and Tollon, C. (2010). The origin and evolution of a recent agricultural weed: population genetic diversity of weedy populations of sunflower (*Helianthus annuus L.*) in Spain and France. *Evolutionary Applications*, 4(3):499–514.

Müller-Schärer, H., Schaffner, U., and Steinger, T. (2004). Evolution in invasive plants: implications for biological control. *Trends in Ecology & Evolution*, 19(8):417–422.

Muminović, J., Merz, A., Melchinger, A. E., and Lübberstedt, T. (2005). Genetic structure and diversity among radish varieties as inferred from AFLP and ISSR analyses. *Journal of the American Society for Horticultural Science*, 130(1):79–87.

Neuwirth, E. (2014). ColorBrewer Palettes [R package RColorBrewer version 1.1-2].

Neve, P., Vila-Aiub, M., and Roux, F. (2009). Evolutionary-thinking in agricultural weed management. *New Phytologist*, 184(4):783–793.

Novy, A., Flory, S. L., and Hartman, J. M. (2012). Evidence for rapid evolution of phenology in an invasive grass. *Journal of Evolutionary Biology*, 26(2):443–450.

Palacio-López, K., Beckage, B., Scheiner, S., and Molofsky, J. (2015). The ubiquity of phenotypic plasticity in plants: a synthesis. *Ecology and Evolution*, 5(16):3389–3400.

Palumbi, S. R. (2001). Evolution - Humans as the world's greatest evolutionary force. *Science*, 293(5536):1786–1790.

Panetsos, C. and Baker, H. G. (1967). The origin of variation in "wild" *Raphanus sativus* (Cruciferae) in California. *Genetica*, 38:243–274.

Patterson, N., Price, A. L., and Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genetics*, 2(12):2074–2093.

Pickersgill, B. (1971). Relationships Between Weedy and Cultivated Forms in Some Species of Chili Peppers (Genus *Capsicum*). *Evolution*, 25(4):683–691.

Pritchard, J. and Stephens, M. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155:945–959.

Pritchard, J. K., Wen, X., and Falush, D. (2010). Documentation for STRUCTURE software: Version 2.3. University of Chicago.

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, 16(3):608–627.

R Core Team (2015). R: A Language and Environment for Statistical Computing.

Reznick, D. N. and Ghalambor, C. K. (2001). The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica*, 112-113:183–198.

Richards, C. L., Bossdorf, O., Muth, N. Z., Gurevitch, J., and Pigliucci, M. (2006). Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. *Ecology Letters*, 9(8):981–993.

Ridley, C. E. and Ellstrand, N. C. (2008). Evolution of enhanced reproduction in the hybrid-derived invasive, California wild radish (*Raphanus sativus*). *Biological Invasions*, 11(10):2251–2264.

Ridley, C. E. and Ellstrand, N. C. (2010). Rapid evolution of morphology and adaptive life history in the invasive California wild radish (*Raphanus sativus*) and the implications for management. *Evolutionary Applications*, 3(1):64–76.

Ridley, C. E., Kim, S. C., and Ellstrand, N. C. (2008). Bidirectional history of hybridization in California wild radish, *Raphanus sativus* (Brassicaceae), as revealed by chloroplast DNA. *American Journal of Botany*, 95(11):1437–1442.

Rius, M. and Darling, J. A. (2014). How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution*, 29(4):233–242.

Sahli, H. F., Conner, J. K., Shaw, F. H., Howe, S., and Lale, A. (2008). Adaptive Differentiation of Quantitative Traits in the Globally Distributed Weed, Wild Radish (*Raphanus raphanistrum*). *Genetics*, 180(2):945–955.

Sakai, A. K., Allendorf, F. W., Holt, J. S., and Lodge, D. M. (2001). The Population Biology of Invasive Species. *Annual Review of Plant Biology*, 32:305–332.

Schroeder, J. (1989). Wild Radish (*Raphanus raphanistrum*) Control in Soft Red Winter Wheat (*Triticum aestivum*). *Weed Science*, 37(1):112–116.

Shen, D., Sun, H., Huang, M., Zheng, Y., Qiu, Y., Li, X., and Fei, Z. (2013). Comprehensive analysis of expressed sequence tags from cultivated and wild radish (*Raphanus spp.*). *BMC Genomics,* 14(1):72.

Snow, A. A. and Campbell, L. G. (2005). Can feral radishes become weeds? In *Crop Ferality and Volunteerism,* pages 193–208. CRC Press.

Snow, A. A., Culley, T. M., Campbell, L. G., Sweeney, P. M., Hegde, S. G., and Ellstrand, N. C. (2010). Long-term persistence of crop alleles in weedy populations of wild radish (*Raphanus raphanistrum*). *New Phytologist*, 186(2):537–548.

Snow, A. A., Uthus, K. L., and Culley, T. M. (2001). Fitness of Hybrids Between Weedy and Cultivated Radish: Implications for Weed Evolution. *Ecological Applications*, 11(3):934–943.

Stanton, M. L., Snow, A. A., and Handel, S. N. (1986). Floral evolution: attractiveness to pollinators increases male fitness. *Science*, 232(4758):1625–1627.

Stewart Jr, C. N., Tranel, P. J., Horvath, D. P., Anderson, J. V., Rieseberg, L. H., Westwood, J. H., Mallory-Smith, C. A., Zapiola, M. L., and Dlugosch, K. M. (2009). Evolution of Weediness and Invasiveness: Charting the Course for Weed Genomics. *Weed Science*, 57(5):451–462.

Thurber, C. S., Reagon, M., Olsen, K. M., Jia, Y., and Caicedo, A. L. (2014). The evolution of flowering strategies in US weedy rice. *American Journal of Botany*, 101(10):1737–1747.

Vigueira, C. C., Olsen, K. M., and Caicedo, A. L. (2012). The red queen in the corn: agricultural weeds as models of rapid adaptive evolution. *Heredity*, 110:303–311.

Wang, N., Kitamoto, N., Ohsawa, R., and Fujimura, T. (2008). Genetic diversity of radish (*Raphanus sativus*) germplasms and relationships among worldwide accessions analyzed with AFLP markers. *Breeding Science*, 58(2):107–112.

Warwick, S. I. and Francis, A. (2005). The biology of Canadian weeds. 132. *Raphanus raphanistrum L. Canadian Journal of Plant Science*, 85(3):709–733.

Webster, T. M. and Macdonald, G. E. (2001). A Survey of Weeds in Various Crops in Georgia. *Weed Technology*, 15(4):771–790.

Whitney, K. D. and Gering, E. (2015). Five decades of invasion genetics. *New Phytologist,* 205(2):472–475.

Yamagishi, H. and Terachi, T. (2003). Multiple origins of cultivated radishes as evidenced by a comparison of the structural variations in mitochondrial DNA of *Raphanus. Genome*, 46(1):89–94.

Yamane, K., Lü, N., and Ohnishi, O. (2005). Chloroplast DNA variations of cultivated radish and its wild relatives. *Plant Science*, 168(3):627–634.

Yamane, K., Lü, N., and Ohnishi, O. (2009). Multiple origins and high genetic diversity of cultivated radish inferred from polymorphism in chloroplast simple sequence repeats. *Breeding Science,* 59:55–65.

Yoon, M. S., Lee, J., Kim, C. Y., and Baek, H. J. (2007). Genetic relationships among cultivated and wild *Vigna angularis* (Willd.) Ohwi et Ohashi and relatives from Korea based on AFLP markers. *Genetic Resources and Crop Evolution*, 54(4):875–883.

Ziffer-Berger, J., Hanin, N., Fogel, T., Mummenhoff, K., and Barazani, O. (2014). Molecular phylogeny indicates polyphyly in *Raphanus L.* (Brassicaceae). *Edinburgh Journal of Botany*, 72(01):1–11.

Zizumbo-Villarreal, D., Colunga-GarcíaMarín, P., de la Cruz, E. P., Delgado-Valerio, P., and Gepts, P. (2005). Population Structure and Evolutionary Dynamics of Wild–Weedy–Domesticated Complexes of Common Bean in a Mesoamerican Region. *Crop Science*, 45(3):1073–11.

**CHAPTER II: Signatures of selection in weedy radish**

## Introduction

Agricultural weeds cost the United States more than 34 billion a year (Pimentel et al., 2005), mostly from production losses. However, a small number of weed lineages - twelve plant families and fewer than 200 species - account for the majority of these losses (Holm, 1978,9), which suggests that there are only a few evolutionary routes can lead to 'weediness'. This may be due to the nature of farming itself. In an agricultural setting, practices like tilling create a narrow window for the weed life cycle, excluding plants requiring more than a few months to go from germination to setting seed. This cyclic and intense disturbance of agricultural fields is unique in nature and exerts a strong selective force on both crops and weeds. Research on agricultural weeds shares many of the same advantages as studying the crops themselves: recent origin (~12,000 years), economic importance and availability of historical data (Vigueira et al., 2012). However, weeds have evolved without directed human selection (Vigueira et al., 2012; Warwick and Stewart, 2005), and so can additionally inform us about the processes of natural selection in an unnatural environment. Understanding why these particular plants transitioned to become weeds, and how they evolved to exist in such a regimented ecosystem may provide insights into the process of natural selection and could guide improvements in weed management (Vigueira et al., 2012).

Weedy plants are often characterized by their ability to colonize disturbed habitats, and such colonization events are often associated with rapid evolution (Reznick and Ghalambor, 2001), whether the disturbance is natural or caused by human activity (de Wet and Harlan, 1975). Studies have found evidence for both de novo resistance, as well as rapid selection for traits like acetolactate synthase (ALS) resistance, in response to repeated herbicide use (Barrett, 1983; Busi et al., 2013; Derksen et al., 2002; Jasieniuk et al., 1996; Neve et al., 2009; Tranel and Wright,

2002; Vigueira et al., 2012). Similarly, it is proposed that vegetative mimicry in weedy sorghum, teosinte, and rices is due to un-intentional human selection during hand-weeding, resulting in seedling traits that closely resemble important crops (Barrett, 1983). While traits like herbicide resistance and crop mimicry evolve rapidly, and clearly are important for weed survival in a modern agricultural field, they were unlikely to be important for initial colonization. For these traits to be selected for, weeds must already be growing in an agricultural field and exposed to human interventions (Barrett, 1983; C Neal Stewart, 2009; Tranel and Wright, 2002; Vigueira et al., 2012).

Instead, due to the nature of most farming practices, we might expect the strongest initial selection to be for reduced flowering time, increased rate of growth, and earlier seed set (Barrett, 1983; Campbell and Snow, 2009; Snow and Campbell, 2005; Warwick and Stewart, 2005). While these traits are often cited as markers of weediness (Baker, 1965; Ellstrand et al., 2010; Neve et al., 2009; Sutherland, 2004; Vigueira et al., 2012; Warwick and Stewart, 2005), studies of the genetics of these key phenotypic adaptations to agriculture are mostly lacking (Ellstrand et al., 2010; Vigueira et al., 2012). The adaptive evolution of these other weedy traits, particularly rapid seed set and loss of any over-wintering requirement is probably an important indicator of early weed potential, yet we have little information about genes influencing these traits outside of Arabidopsis thaliana, which is a poor model for weeds (Basu et al., 2004).

The lack of detailed genetic studies of these traits is likely due to the surprising complexity of weed origins. There are three potential origin routes for agricultural weeds: wild populations invading fields, crops going feral, and crop/wild hybridization (de Wet, 1966; de Wet and Harlan, 1975; Vigueira et al., 2012). Weed species with varying origins may therefore take

51

different evolutionary trajectories to arrive at the same set of weed traits. Wild plants, for example, may be initially be under selection for characters that improve their ability to survive in crop fields such as faster growth. However, feral crops will have already undergone many generations of artificial selection for these traits, and may already be resistant to herbicides, and so might be under strong selection for seed shattering and other escape traits (Vigueira et al., 2012). Hybrids which might inherit both agricultural growth habits and wild reproductive traits, may require little or no selection to successfully invade. Understanding both the origins and early adaptations of lineages with high concentrations of weedy species may not only help improve control methods (McNeill, 1976; Müller-Schärer et al., 2004; Neve et al., 2009; Stewart Jr et al., 2009; Vigueira et al., 2012) for current problem species, but may also provide insight into approaches for preventing future weeds (Higgins et al., 1978).

One emerging model for studying both rapid evolution and weed evolution is weedy radish, *Raphanus raphanistrum ssp. raphanistrum*. Weedy radish is an ideal weed for answering questions about weed lineages as it is both a typical agricultural weed and economically important. Weedy radish is a member of Brassicaceae, one of the twelve major agricultural weed families, and is considered one of the world's worst weeds (Holm, 1997) and is the worst dicot weed in Australia (Warwick and Francis, 2005). Weedy radish primarily invades small grain fields throughout the world (Warwick and Francis, 2005), and in the last 200 years weedy *R. r. raphanistrum* has spread to every continent except Antarctica (Holm, 1997).

Weedy radish belongs to a small family of plants with wild, crop and weedy members. *R. sativus*, crop radish, has been domesticated into four major cultivar types. Two, Daikons and Europeans, are common cooking radishs, while Oilseeds and Rattails have been bred mainly as

seed crops. In addition to *R. sativus*, weedy radish also has two wild (non-weedy) relatives, *R. r. ssp. landra*, a native of the Mediterranean commonly found on dunes; and *R. pugioniformis*, which is found in the eastern Mediterranean but is largely unstudied. In addition, a number of genomic tools are available for weedy radish and its close relatives, including draft genomes of both weedy *R. r. raphanistrum* (Moghe et al., 2014) and several crop cultivars (Jeong et al., 2016; Kitashiba et al., 2014; Mitsui et al., 2015).

Perhaps not surprisingly, previous work suggests that weedy radish is most closely related to native *R.r. raphanistrum* (Chapter I; Shen et al., 2013b). However, weedy and native *R.r. raphanistrum* are phenotypically distinct for traits likely to be important for a native turned weed. For example, weedy radish flowers twenty-four to fifty-eight days earlier than native *R.r. raphanistrum* (Chapter I) and spends fewer resources on basal leaf development (Sahli et al., 2008).

We have used RAD-TAG, a reduced representation sequencing strategy, to confirm and expand our previous findings in a subset of populations from Chapter I. In this study, we have increased the number of genome wide SNPs by an order of magnitude and have used these markers to both access the interrelatedness of the *Raphanus* clade, and to find loci under selection in the weeds. Weedy radish is presumed to have evolved sometime in the last ~12,000 years in conjunction with the advent of farming (Vigueira et al., 2012), and are quite phenotypically divergent from native radish. As twelve thousand years is short in evolutionary terms, we expect selective events from this time period to be detectable as reduced genetic diversity at selected loci (Smith and Haigh, 1974). Further, we would expect that at least some of the genetic differences between

closely related populations are likely to underlie phenotypic divergence between those populations (Nielsen, 2005).

## Methods

### Populations

To obtain a large number of markers for selection analysis and validating *Raphanus* we did single end, reduced representation, Genotype-by-Sequencing (GBS) experiment (Elshire et al., 2011) with a total of 259 individuals. We included representatives from all named species and sub-species of *Raphanus* and all four cultivated varieties (Table 2-1.) For weeds and natives, we sequenced 16 individuals per population. As we expected crop populations to be less genetically variable than the non-domesticated varieties, we sequenced fewer (5-7) crop individuals per variety. This allowed us to sample more native populations than would have been possible with equal sample sizes. Both the library prep with the pstI restriction enzyme, and sequencing with an Illumina HiSeq were performed by the BRC Genomic Diversity Facility at Cornell.

### Mapping and quantification

We processed the raw reads using the process radtags function of STACKS (Catchen et al., 2013,0) to deconvolute our samples and remove the sequencing barcodes and adaptors.

We used GSNAP (Wu and Watanabe, 2005) to align reads to the published draft genome for *R.r. raphanistrum* (Moghe et al., 2014) for use in the pSTACKS pipeline. Due to restrictions imposed by our downstream STACKS analysis, we mapped using a kmer size of 15, did not allow

**Table 2-1: Populations.** Variety/Range refers to whether a wild population was collected inside (Native Range) or outside the Mediterranean region; or to the convariety name for crops. Source refers to the original location of source populations in the case of wild populations, or the company that cultivar seed was purchased from. Flowering time is the average days to flower for that population. % Vernalization is the average percentage of plants that require cold treatment to flower. Whenever possible, Flowering time and vernalization estimates are from Figure 1-3. RA226 and RA444 flowering times were obtained from a small unpublished experiment in the Conner lab, each with (N=20). RA444 grew, but never bolted, and presumably has an absolute vernalization requirement. We have no flowering information for RA808 or YEIL.

| Population | N | Variety/Range | Source | Flowering Time | % Vernalization |
|---|---|---|---|---|---|
| ***R.r.landra*** | | | | | |
| PBFR | 16 | Native | France | 336 | 100 |
| RA226 | 16 | Native | Spain | 119 | 100 |
| RA444 | 16 | Native | Spain | NB | 100 |
| RA761 | 16 | Native | Spain | 94 | 88 |
| RA808 | 16 | Native | Spain | | |
| SAES | 16 | Native | Spain | 216 | 81 |
| ***R.r.raphanistrum*** | | | | | |
| AFFR | 16 | Native | France | 40 | 10 |
| DEES | 16 | Native | Spain | 93 | 100 |
| MAES | 16 | Native | Spain | 90 | 61 |
| BINY | 16 | Weedy | New York | 31 | 10 |
| NAAU | 16 | Weedy | Australia | 26 | 0 |
| ***R.pugioniformis*** | | | | | |
| GMIL | 16 | Native | Israel | 54 | 6 |
| YEIL | 16 | Native | Israel | | |
| ***R.s.convar.sativus*** | | | | | |
| NEJS | 7 | Daikon | John Scheepers | 157 | 77 |
| TOBG | 7 | Daikon | Bountiful Gardens | 173 | 67 |
| ESNK | 7 | European | NK Lawn & Garden | 70 | 9 |
| SPNK | 7 | European | NK Lawn & Garden | 60 | 0 |
| ***R.s.convar.oleifera*** | | | | | |
| AROL | 6 | Rattail | MSU | 47 | 0 |
| OIBG | 5 | Rattail | Bountiful Gardens | 66 | 0 |
| ***R.s.convar.caudatus*** | | | | | |
| RAJS | 6 | Oilseed | John Scheepers | 42 | 0 |
| RABG | 6 | Oilseed | Bountiful Gardens | 43 | 0 |

splicing or terminal alignments, required a minimum coverage of 95%, and kept only those reads with a phred scaled mapping quality of at least 20.

To obtain loci for genus-level comparisons, we used the pSTACKS pipeline (Catchen et al., 2013,0) to align reads between samples and call SNPs. Prior to this analysis, we used the methods described in Paris et al. (2017) to test all reasonable STACKS settings on a subset of this data, and experimentally determined the pSTACKS parameters that maximized our ability to find SNPs without collapsing paralogous loci. As such, we required at least three reads per individual to accept a locus (m=3); and set n, the number of mismatches allowed when building consensus loci within a population to n=7. We used the populations module in STACKS to calculate pairwise Fst and obtain variant lists for various population groupings. Standard summary statistics were computed using the R (R Core Team, 2015) packages 'adegenet' (v.2.0.1) and 'pegas' (v.0.10).

To find loci for signatures of selection, we used the same STACKs pipeline, however we only used *R.r. raphanistrum* individuals to build the catalog and used 'weed' or 'native' as population names.

**SmartPCA**

To compare our findings to that of Chapter I, we used SmartPCA (v.13050 – from the program Eigensoft 6.0.1) (Patterson et al., 2006) to perform an eigendecompo- sition. This program is optimized for genomic markers and rotates the data onto a set of orthogonal axes defined by the amount of variation explained to reveal hidden data structure. As the SmartPCA algorithm is designed for biallelic markers, each SSR marker was expanded into several biallelic markers as

described in Patterson et al. (2006), prior to analysis using a custom script. SmartPCA was designed for human data with a maximum of 23 chromosomes, however the Moghe et al. (2014) genome was assembled into 68331 contigs. As such, we randomly assigned markers to nine pseudo-chromosomes to match the known number of *Raphanus* chromosomes. Multiple iterations of randomization produced exactly the same plot, so although SmartPCA requires this chromosome position information, it doesn't get used in the decomposition, and has no effect on the analysis.

**STRUCTURE**

Similarly, we tested for population structure using the command line version of STRUC- TURE (v2.3.4) (Pritchard and Stephens, 2000). We tested the admixture model for 20 replicates for each potential value of K from 3 to 20. To account for potential ordering effects, each replicate was run with a different randomization of the data input order. We used a burn in of 500,000 cycles followed by one million Markov chain Monte Carlo iterations. We ran the correlated allele frequencies (or F) model, which computes values similar to Fst to model genetic differentiation and does not require genetic linkage information. Our populations were not constrained to a single Fst, and alpha (the admixture parameter) was inferred for the overall dataset as suggested in the manual (v. 2.3) (Pritchard et al., 2010). We used the default settings for all priors.

From this, we calculated optimal values for K using both the mean log probability of K for each iteration as in Pritchard and Stephens (2000) and the largest delta K as described in Evanno et al. (2005). These calculations were run using the web-based version of Structure Harvester (Web v0.6.94 July 2014, Plot vA.1 November 2012, Core vA.2 July 2014) (Earl and vonHoldt, 2012).

All SmartPCA and STRUCTURE analyses were run on the HPCC at Michigan State University. We plotted results from SmartPCA and STRUCTURE analysis using custom bash and R scripts (R Core Team, 2015).

**Fst outlier detection**

We used PGDSpider version 2.1.1.3 (Lischer et al.) to convert vcf output from the populations module of STACKS into input files for BayeScan version 2.1 (Ecology and 2012). To improve our ability to detect outliers, we only included loci with a minor allele frequency greater than .05, used an odds ratio of 3, and made several data subsets. As we expected weedy *R.r. raphanistrum* to be most closely related to natives of the same subspecies, we did a paired analysis of each weed (BINY and NAAU) against each native *R.r. raphanistrum* (AFFR, DEES and MAES). We also analyzed a subset where the weeds were treated as a single group and compared to native *R.r. raphanistrum*, as a group. To look for genes of interest among Fst outliers, we blasted whole contigs using NCBI megaBLASTN 2.8.0+ against the nucleotide collection.

# Results

## Sequencing, Mapping and Quantification

We obtained approximately two-hundred million reads per lane, for an average of 2.1 million reads per individual. While read quality estimates from the Illumina HiSeq quality software were reported to be within normal range, we found extensive contamination primer and adapter contamination. The contamination appears to be due to an obsolete paired-end PCR 2.0 primer, and/or the matching paired end sequencing adapter. This sequence always occurs as the reverse complement of the actual primer/adapter sequence, and always on the 3' end of the read.

Approximately half of all reads have at least 35 bases of contamination, and approximately ten

percent of those (5% of all reads) are contaminated with the full 62 base primer sequence. As

STACKS cannot process variable length reads, all contaminated sequences were lost.
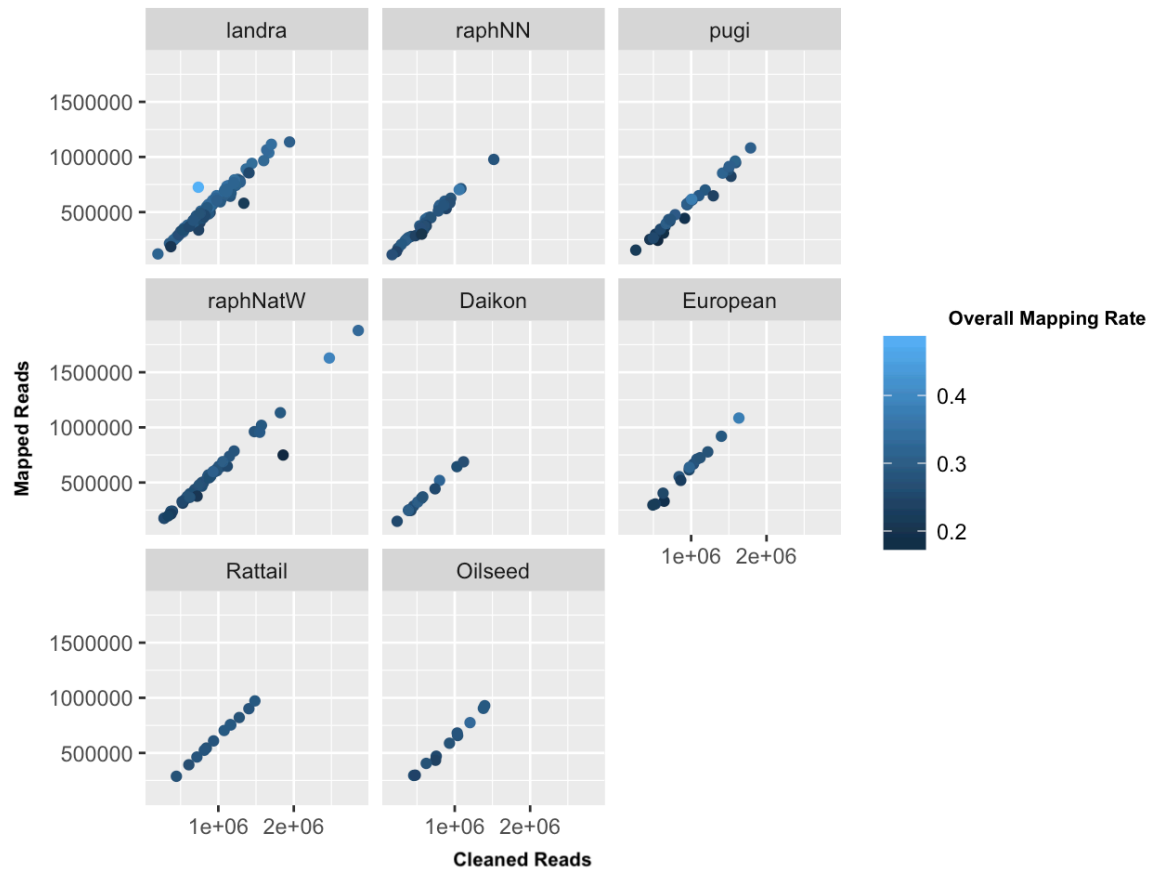


**Figure 2-1: Comparison of mapping rates per individual.** The x-axis shows the number of reads available to GSNAP for mapping. The y-axis shows the number of reads successfully mapped. Points are colored by overall mapping rate: Number of reads mapped Number of reads from sequencer. Reads are split by group to show potential sequencing bias.

Approximately half of the remaining reads mapped successfully for an overall mapping rate of

~25% for most individuals (Fig. 2-1). Although we mapped to the weedy *R.r. raphanistrum*

genome, we saw no significant difference in mapping rates from other taxa.

STACKS built an average of 6526 loci per individual with a mean depth of 79.4 reads. However, these numbers were quite variable, with an average maximum depth of 3508 reads, and some depths reaching as high as 17790.

For the overall dataset used for genus-level analysis, we found 70888 SNPs in at least one heterozygote. These comprised 26686 alleles across all samples. After filtering to accept only SNPs which were present in at least 20 of our 21 populations and in at least 80% of individuals of each population we were left with 1069 SNPs available for downstream analysis.

For the dataset with just *R.r. raphanistrum*, we found 17126 alleles from 41188 SNPs. After filtering for SNPs that were in at least 80% of individuals for both the weeds and natives, we were left with 1891 SNPs for downstream analysis.

Although we purchased, and the sequencer output is consistent with, single end sequencing runs, the final SNPs appear to be from paired end sequencing. Many markers, appear in pairs with zero to a hundred bases between them. Alignments for these markers show that all the reads used to build each marker mapped in opposite orientations as would be expected for paired end reads. As such, the 1069 SNPs in the overall dataset only represent about 600 independent markers. Similarly, in the just *R.r. raphanistrum* dataset ~1000 of the 1891 markers are not very tightly linked.

**Separation by SmartPCA consistent with native *R.r. raphanistrum* as weed origin**

In our SmartPCA analysis, all individuals are tightly clustered by population, and except for *R.r. landra*, and perhaps *R. pugiformis*, taxa also cluster into groups; all four cultivars of *R. sativus* also cluster as a single unit (Fig. 2-2). The first two principle components of our SmartPCA show

strong differentiation between *R.r. raphanistrum* as a whole and all other *Raphanus* taxa (Fig. 2-2, top). Similarly, *R.r. raphanistrum* have some separation from the other taxa in PCs three and four, however the magnitude of difference is somewhat smaller, and MAES, while differentiated from the other sub-species, does not cluster with the rest of *R.r. raphanistrum* (Fig. 2-2, bottom). Within *R.r. raphanistrum*, the weedy plants (orange) are intermediate between native *R.r. raphanistrum, R.r. landra* and *R. sativus* in all of the first four PCs.

**Clustering analyses improves with reduced SNP set**

We found little genetic structure when clustering populations by pairwise-Fst using all 1069 markers (Fig. 2-3, top). Filtering out SNPs with a minor allele frequency less than .05 resulted in clustering that better reflects the groupings suggested by SmartPCA. (Fig. 2-3, bottom). This filtering step reduces the number of markers to 331 (~265 independent markers). Although the reduced marker set has more ability to cluster populations, it still has many markers with low minor allele frequencies, and most have very little variation between populations (Fig. 2-4).
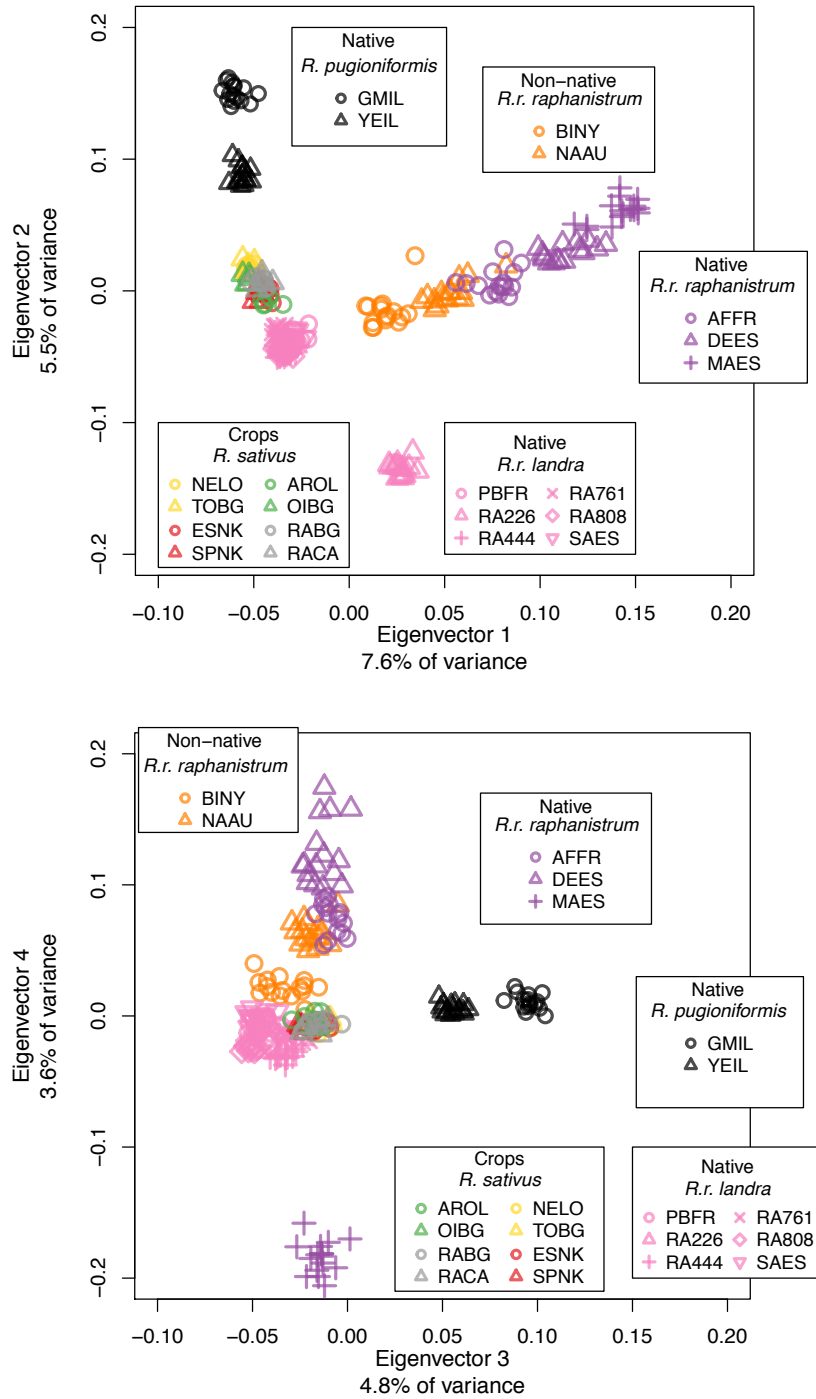
**Figure 2-2: Principal components analysis plots for twenty-one populations of *Raphanus* at 1069 genome-wide SNPs.** Each point is an individual, and each population is represented by 5 to 16 individuals. Populations are identified by plotting character. Top: The first two eigenvectors. Bottom: Eigenvectors three and four. Together these plots show 21.5% of the variance in these markers.

In STRUCTURE results from runs using all 1069 SNPs, both common estimators of K, the number of distinct groupings, suggest that K=4 is the best fit for this data. However, while K=4 produced the highest likelihoods, replicate runs did not produce similar groupings, and we find at least five distinct and mutually exclusive plots (data not shown). Results from the filtered SNPs give much more consistent STRUCTURE plots across replicates, however the best supported K value is K=16, which is likely spuriously high. As such, we used K=5 based on results from Chapter I which suggest the major divisions are between the three major wild taxa: *R. pugioniformis, R.r. landra*, and *R.r. raphanistrum*; weedy *R.r. raphanistrum*; and *R. sativus*, (Fig. 2-5).

At K=5, we find that the STRUCTURE results largely agree with those of Smart- PCA. *R. pugiformis* and the crops both have their own groups, as do the native *R.r. raphanistrum*. Similarly, *R.r. landra* forms a single group with the exception of RA226, which is relatively far from the rest of the *landra* in PCs one and two (Fig. 2-2). The two weed populations also appear to be admixtures of *R.r. landra* and native *R.r. raphanistrum*, just as they appear in the intermediate zone between these taxa in all four PCs. However, even in these more stable STRUCTURE runs, we still find some replicates with alternative groupings (data not shown).
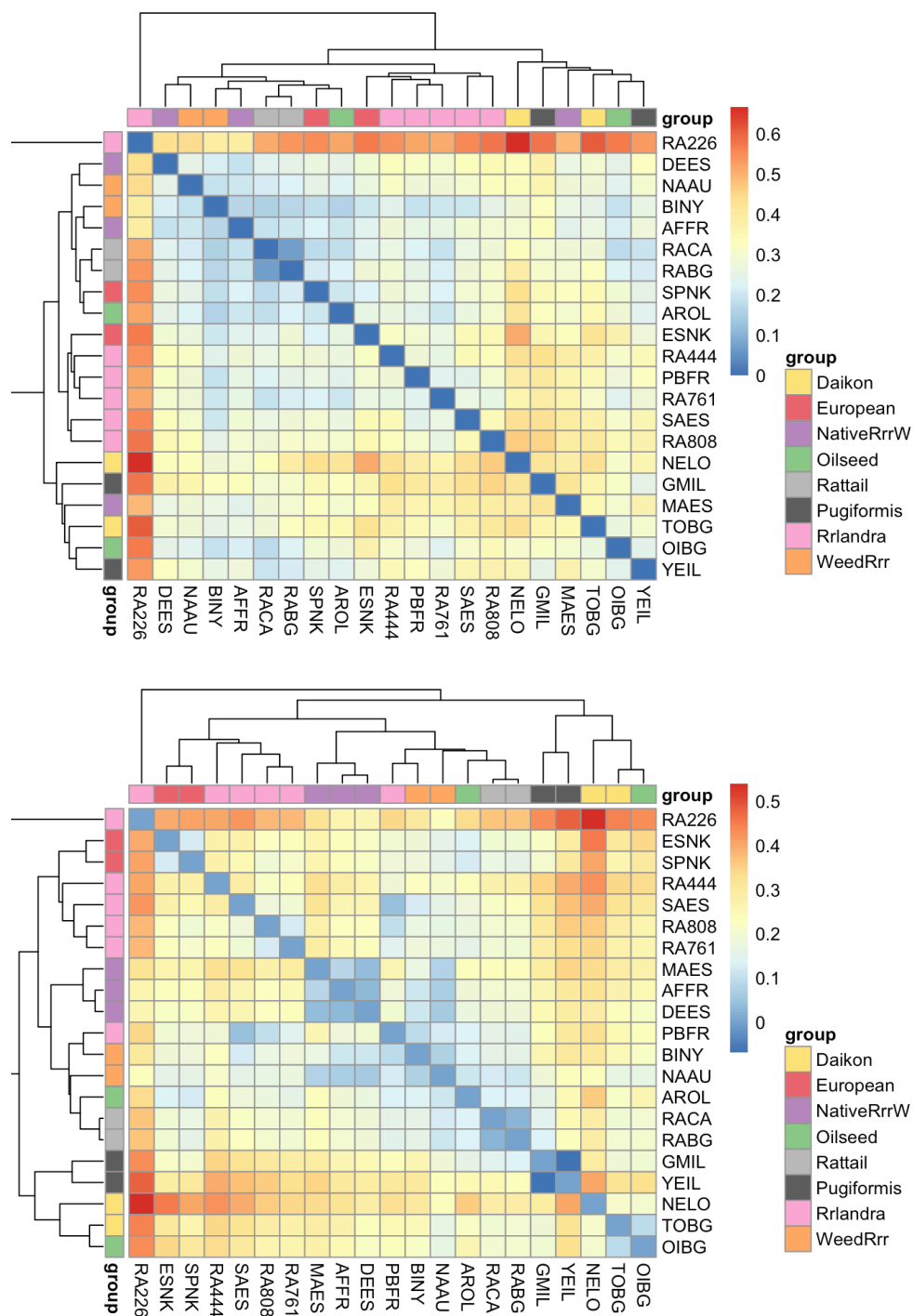
**Figure 2-3: Pairwise Fst clustered by Euclidean distance.** Populations are colored along the axes to match the groups from SmartPCA (Fig. 2-2). For population codes on the other axes, see Table 2-1. Top: Pairwise Fst calculated with all 1069 SNPs shared among all populations. Bottom: Pairwise Fst calculated using the 331 SNPs with a minor allele frequency of at least .05
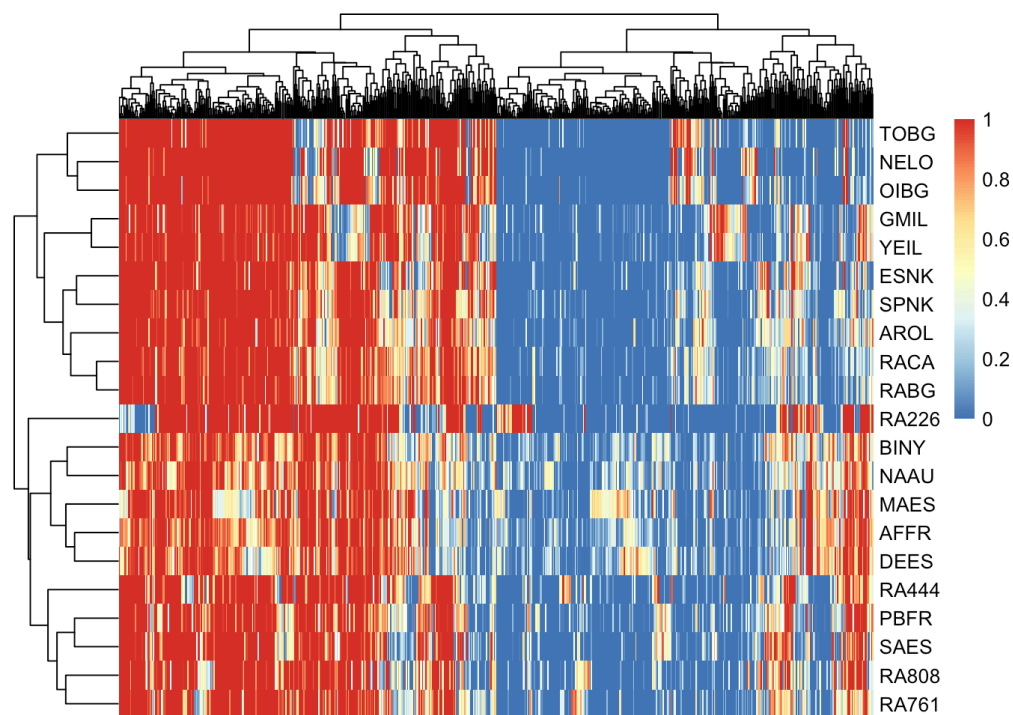
**Figure 2-4: Allele frequencies by population for 331 SNPs with a minor allele frequency of at least .05.** Columns dominated by either red or blue indicate marker with little ability to distinguish between populations.
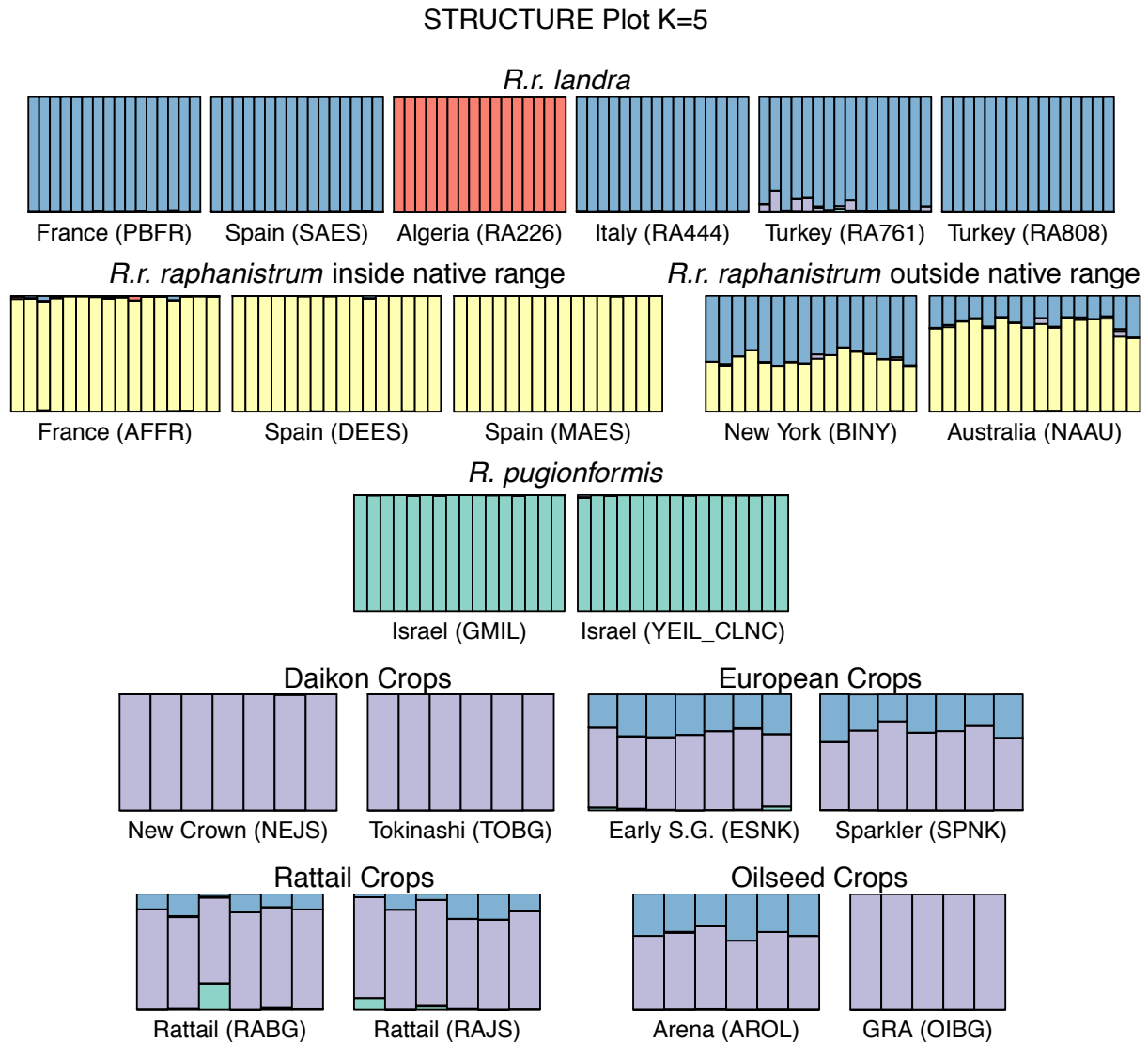
**Figure 2-5: Representative STRUCTURE plot for 331 SNPs with a minor allele frequency of at least .05.** Each large rectangle represents a group of 8-10 individuals, with each individual represented by a vertical bar. A vertical bar of a single color denotes an individual whose genotypes can be entirely attributed to a single background shared by every individual with that color. Individuals with bars of more than one color show evidence for admixture, with the proportion of each genetic background plotted indicated by proportion of each color. Colors are arbitrary.
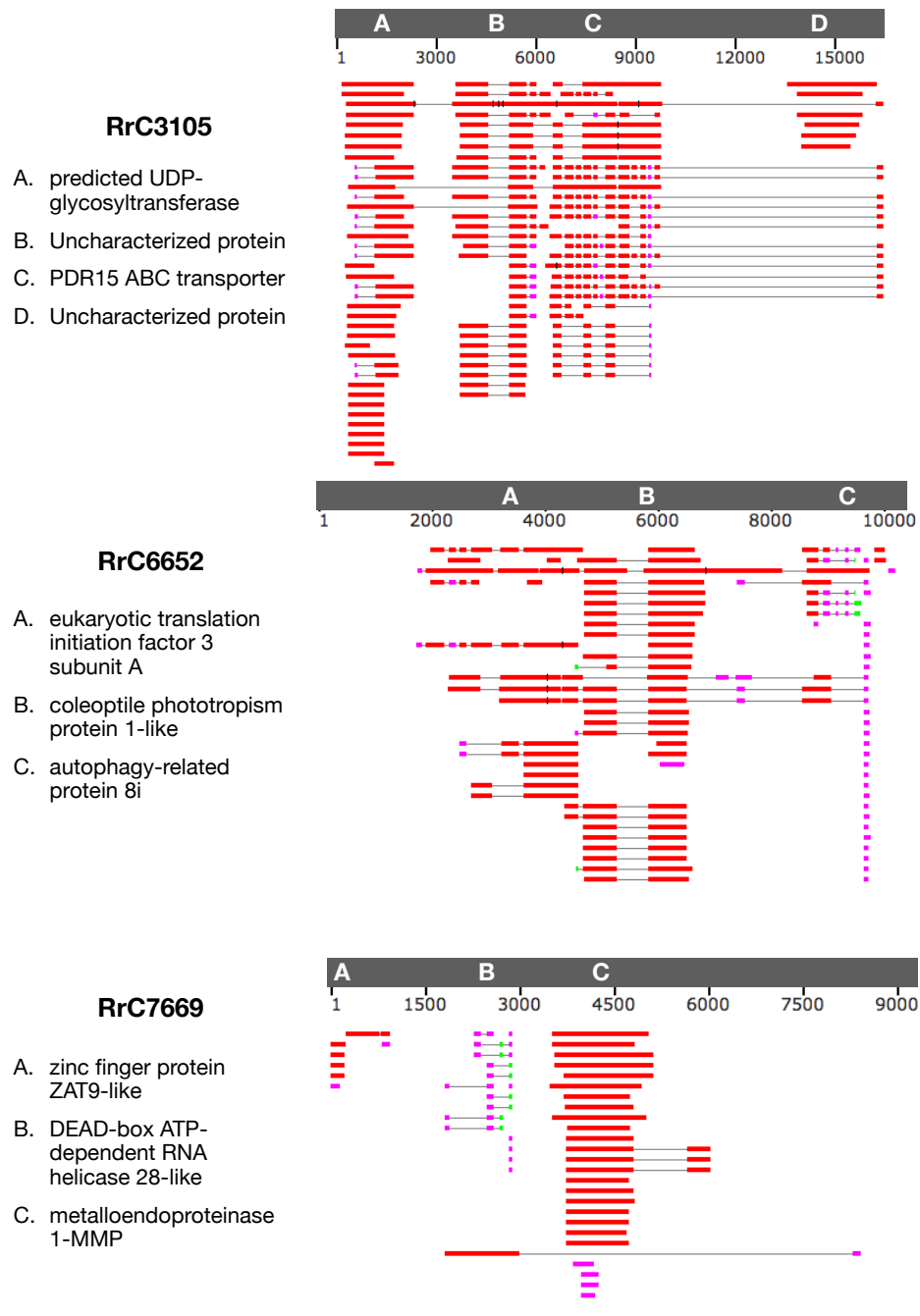
**Figure 2-6: Overview of BLAST hits for Fst outlier contigs.** Only the three contigs with annotated sequences are shown. Letters inside the grey bar representing the genomic region correspond to the annotation listed to the left of the image. Colored bars under each genomic region show sequence matching to each contig sequence, where color represents alignment quality as determined by BLAST. Green is an alignment score of 50-80, purple is 80-200, and red is 200 or greater.

67

**Fst Outliers are from an array of cell processes**

After filtering for markers with a minor allele frequency of at least .05, our *R.r. raphanistrum*

subset contained 949 SNPs, and six of these markers were significant in at least one paired

analysis (Table 2-2). Two of these SNPs were found in multiple comparisons: BINY with AFFR,

and BINY with DEES. However, these two SNPs almost certainly represent a single locus, as the

markers are part of a pair with only ten intervening basepairs.

Altogether, we found five contigs with significant Fst outliers, only three had BLAST hits to

annotated genes (Fig. 2-6). Between these five contigs we find eight candidate genes: a predicted

UDP-glycosyltransferase, a PDR15 ABC transporter, eukaryotic translation initiation factor 3

**Table 2-2: Fst Outliers.** Pops one and two are the populations for each pairwise comparison.
Position is the basepair where the identified SNP is located on the chromosome. Ref (SNP) gives
the reference allele and polymorphism. The qvalue is the FDR adjusted p-value for that outlier.
Comparisons of NAAU with AFFR or MAES had no significant results, nor did the overall
comparison of all weeds with all native *R.r. raphanistrum*

| Pop 1 | Pop 2 | Contig | Position | Ref (SNP) | qvalue |
|-------|-------|--------|----------|-----------|--------|
| BINY | AFFR | RrC3355 | 4370 | G(A) | 0.021 |
| BINY | AFFR | RrC3355 | 4380 | A(G) | 0.021 |
| BINY | DEES | RrC13105 | 5129 | G(A) | 0.024 |
| BINY | DEES | RrC3355 | 4370 | G(A) | 0.026 |
| BINY | DEES | RrC3355 | 4380 | A(G) | 0.025 |
| BINY | DEES | RrC7669 | 1943 | A(T) | 0.024 |
| BINY | MAES | RrC6652 | 4963 | A(G) | 0.049 |
| NAAU | DEES | RrC3105 | 7909 | T(C) | 0.015 |

subunit A, coleoptile phototropism protein 1-like protein, autophagy- related protein 8i, zinc finger protein ZAT9-like, DEAD-box ATP-dependent RNA helicase 28-like, and metalloendoproteinase 1-MMP. The RrC3355 locus, which was significant in two comparisons, had no annotated genes.

## Discussion

We sequenced a diverse collection of populations from across *Raphanus* species to enhance our ability to find population structure, and to find regions that show evidence of selection. Based on previous work (Chapter I), we expected weedy radish (*R.r. raphanistrum*) to be a close relative of native *R.r. raphanistrum* and used an Fst outlier analysis to look for loci that might underlie the large phenotypic differences between these two ecotypes.

### Native *R.r. raphanistrum* as the ancestor to weedy radish

We used three different analyses to access the inter-relatedness of *Raphanus* species. In agreement with previous reports (Chapter I), we find that weedy and native *R.r. raphanistrum* are the most genetically similar to each other among the *Raphanus* as measured by PCA. We found similar results with STRUCTURE, where both are suggestive of a native hybrid origin for the weeds (Figs. 2-2 & 2-5), however a larger number of markers and an explicitly phylogenetic approach would better resolve these relationships.

In keeping with recent previous reports (Chapter I; Shen et al., 2013a), we also find evidence for single origin for crop radish as the cultivars cluster through the first four PCs and in our STRUCTURE analysis. Although there are many earlier reports of multiple crop origins (Kong et al., 2011; Lü et al., 2008; Muminovíc et al., 2005; Wang et al., 2008; Yamagishi and Terachi,

2003; Yamane et al., 2005,0), these reports sampled few, if any, non-crop populations, and likely over-estimated the divergence between cultivars.

Although our results from SmartPCA support previous findings, we found little population structure using pairwise Fsts. Similarly, while the filtered marker STRUCTURE analysis largely agree with the SmartPCA, and out-performed the version using all possible SNPs, about half of our filtered STRUCTURE replicates still had other population groupings. These differences in our ability to see population structure are likely due to how these methods deal with the relatively low minor allele frequencies in our markers and the number that appear in high LD pairs. One of the underlying assumptions of the STRUCTURE algorithm is that all of the SNPs are independent, however many of our loci are very tightly linked, which may be why our groupings are unstable. In the two alternative STRUCTURE scenarios, either PBFR and SAES were split from *R.r. landra* and *R. pugionformis* was lumped with the crops; or PBFR and SAES were split from *R.r. landra* and RA226 was clustered with native *R.r. raphanistrum*. Neither of these alternatives is a particularly good fit for the PCA results, however analysis with more markers with low LD would improve these results. Similarly, the pairwise Fst comparison is overwhelmed by mostly uninformative loci where the minor allele is present in only a handful of individuals across the entire genus. Re-running these analyses without RA226, which is clearly a genetic outlier, may improve the clustering results.

In contrast, SmartPCA uses an eigendecomposition to find the axes with the most variation, and so is not overwhelmed by uninformative sites, and is not greatly impacted by linkage. As such, our SmartPCA results are likely a more reliable indicator of population structure, however it's

possible that the loci driving the differentiation are not neutral, and we are seeing some evidence of selection rather than background diversity.

Interestingly, in the first eigenvector the *R.r. raphanistrum* populations are arranged in a line roughly in order of flowering phenotype, with the slow flowering MAES on the right, and weeds on the left (Fig. 2-2, top). As the first eigenvector is the axis of most variation in the dataset, and the *R.r. raphanistrum* populations account for most of this axis, it is possible that the PCA was biased by genes under selection. While we cannot entirely rule that out, the PCA loadings indicate that seven SNPs are the primary drivers of differentiation in this axis, and their corresponding contigs do not BLAST to genes annotated for flowering time genes. In most cases, the contigs are only annotated with unknown protein predictions. As flowering time genes are well annotated in *Arabidopisis*, this suggests that the PCA is reflecting neutral variation and is a reliable indicator of population structure.

**Library preparation problems caused a severe reduction in power**

We used a Genotype-By-Sequencing (GBS) approach to obtain a large number of markers to validate and extend the results of Chapter I, however our data suffered from low overall quality. Approximately half of all reads were contaminated by either a paired- end adapter or paired end PCR primer from an obsolete Illumina kit. This contamination was always in the reverse compliment at the 3' end of the read which suggests that the DNA used for the sequencing prep was either very degraded or was over-digested such that we got extensive read-through of the reverse adapter. However, while some samples in the study were from relatively old frozen DNA, age of sample doesn't seem to correlate with either number of reads obtained from the sequencer or mapping ability (Fig. 2-7). As these single end reads are contaminated with paired

end adapters, and our marker distribution suggests they behave as paired end reads, it seems

likely that there were previously un-identified problems with the library preparation. This

resulted in a loss of most of the potential information from the sequencing data.
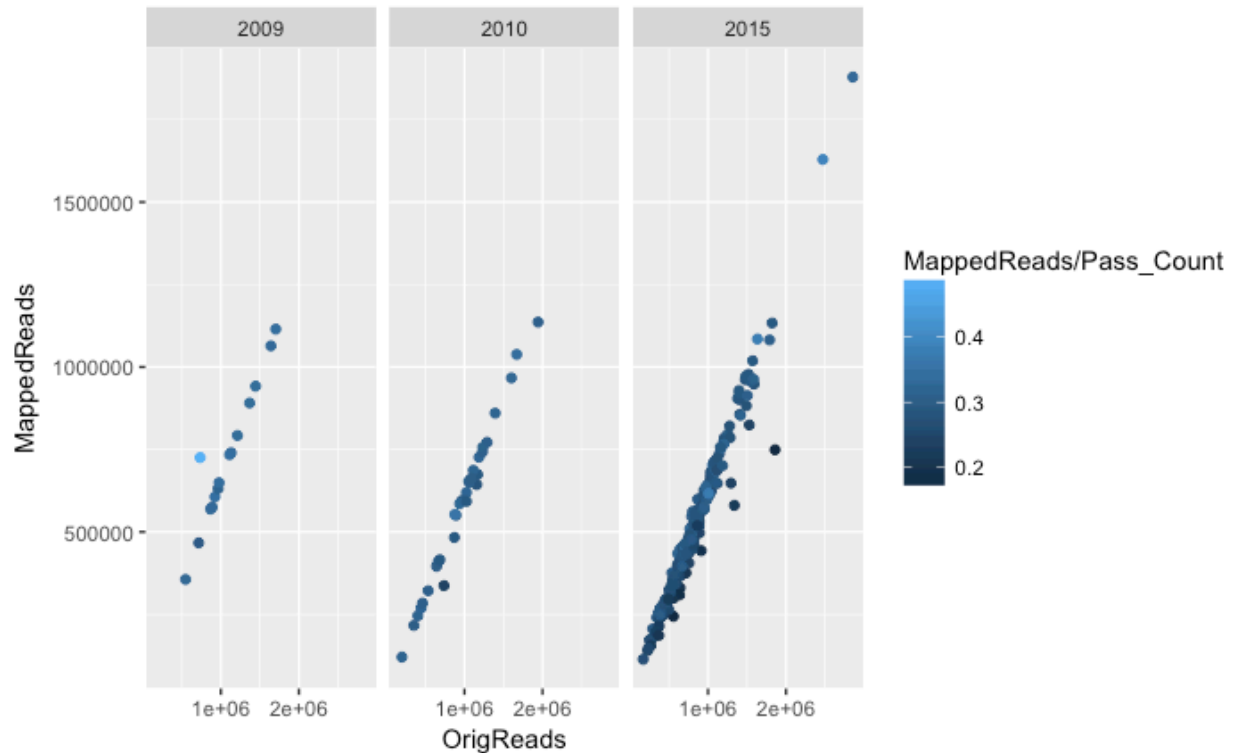


**Figure 2-7: Comparison of mapping rates per individual separated by date of DNA preparation.** The x-axis shows the number of reads available to GSNAP for mapping. The y-axis shows the number of reads successfully mapped. Points are colored by overall mapping rate: Number of reads mapped Number of reads from sequencer. Reads are split by date to show potential sequencing bias.

We suffered a further loss of power due to the majority of our potential markers having

extremely low minor allele frequencies. In many of the SNPs that were dropped, the minor allele

frequency was 1/516, as the minor allele was found in only a single heterozygote. Even after

filtering, many markers showed only minor differences in allele frequency between populations

(Fig. 2-4). This is surprising, as *Raphanus* is an obligate outcrosser and these same populations

were previously found to be quite heterozygous at neutral markers (Chapter I). This effect does not seem to be due to the analysis pipeline, as changing the number of mismatches allowed in mapping, or the assembly parameters in STACKS had almost no effect on allele frequencies (data not shown).

Together, these issues resulted in a large overall reduction in our ability to detect selection. Genome wide searches for changes in Fst typically require many thousands of markers to cover the genome densely enough that some happen to be in linkage disequilibrium (LD) with the actual locus under selection. Sparse markers are unlikely to be in LD with loci of interest and are therefore unlikely to catch any but the newest artifacts of directional selection. Low numbers of markers combined with low levels of allele diversity at those SNPs made it almost impossible to detect loci with our Bayesian method. However low power is a relatively common problem with reduced representation studies Lowry et al. (2016), so simply increasing the quality of the data may not improve our ability to find selective sweeps. Re-running the analysis in a non-Bayesian framework where candidate alleles don't have to overcome the prior distribution might enhance our ability to find outlier loci, at the cost of an increased false positive rate.

**Outlier contigs suggest changes in protein production**

Although we had little power, our Fst outlier analysis identified five regions with significant pairwise deviations. Two of these regions had no significant BLAST results, however we found eight annotated genes among the other three regions. These genes are involved mainly in transcription, protein production and stress response. Although it is difficult to directly relate these genes to flowering time, they do seem to be reasonable candidates for morphological changes. For example, we find genes involved in regulation at all levels: UDP-

glycosyltransferase, is part of a pathway that regulates hormone metabolism, zinc finger protein ZAT9 is a transcription factor, this DEAD-box protein is involved in RNA splicing, and coleoptile phototropism protein 1 mediates protein ubiquitination. Future studies to determine which, if any, of these genes are the actual locus under selection might shed light on the mechanisms that underlie differences between weedy and native *R.r. raphanistrum*.

**REFERENCES**

# REFERENCES

H. G. Baker. Characteristics and modes of origin of weeds. In H. G. Baker and G. L. Stebbins, editors, The genetics of colonizing species, pages 147–172. *cabdirect.org*, 1965.

S. C. H. Barrett. Crop mimicry in weeds. Economic Botany, 37(3):255–282, 1983.

C. C. Basu, M. D. M. Halfhill, T. C. T. Mueller, and C. N. C. Stewart. Weed genomics: new tools to understand weed biology. *Trends in plant science*, 9(8):391–398, Aug. 2004.

R. Busi, P. Neve, and S. Powles. Evolved polygenic herbicide resistance in *Lolium rigidum* by low-dose herbicide selection within standing genetic variation. *Evolutionary Applications,* 6(2):231–242, Feb. 2013.

J. C Neal Stewart. Weedy and Invasive Plant Genomics. *Wiley-Blackwell*, Sept. 2009.

L. G. Campbell and A. A. Snow. Can feral weeds evolve from cultivated radish (*Raphanus sativus*, Brassicaceae)? *American Journal of Botany*, 96(2):498–506, Feb. 2009.

J. Catchen, P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11):3124–3140, May 2013.

J. M. Catchen, A. Amores, and P. Hohenlohe. Stacks: building and genotyping loci de novo from short-read sequences. *G3 Genes Genomes Genetics*, 2011.

J. de Wet. The Origin of Weediness in Plants. *Proc. of the Okla. Acad. of Sci*., pages 14–17, 1966.

J. de Wet and J. R. Harlan. Weeds and domesticates: evolution in the man-made habitat. *Economic Botany*, 29(2):99–107, 1975.

D. A. Derksen, R. L. Anderson, R. E. Blackshaw, and B. Maxwell. Weed dynamics and management strategies for cropping systems in the northern Great Plains. *Agronomy Journal*, 94(2):174–185, 2002.

D. A. Earl and B. M. vonHoldt. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2):359–361, June 2012.

M. F. Ecology and 2012. BayeScan v2. 1 user manual. cmpg.unibe.ch.

N. C. Ellstrand, S. M. Heredia, J. A. Leak-Garcia, J. M. Heraty, J. C. Burger, L. Yao, S. Nohzadeh-Malakshah, and C. E. Ridley. Crops gone wild: evolution of weeds and invasives from domesticated ancestors. *Evolutionary Applications*, 3(5-6):494–504, July 2010.

R. J. Elshire, J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS One*, 6(5):e19379, May 2011.

G. Evanno, S. Regnaut, and J. Goudet. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, 14(8):2611–2620, July 2005.

R. Higgins, E. Heikes, and H. Kempen. The Case for Preventitive Weed Management. *Proceedings of the Western Society of Weed Science*, 31:35–40, 1978.

L. Holm. Some characteristics of weed problems in two worlds. *Proceedings of the Western Society of Weed Science*, 31:3–12, 1978.

L. G. Holm. World Weeds. Natural Histories and Distribution. *John Wiley & Sons Inc*, Mar. 1997.

M. Jasieniuk, A. L. Brûlé-Babel, and I. N. Morrison. The evolution and genetics of herbicide resistance in weeds. *Weed Science*, 44(1):176–193, 1996.

Y.-M. Jeong, N. Kim, B. O. Ahn, M. Oh, W.-H. Chung, H. Chung, S. Jeong, K.-B. Lim, Y. J. Hwang, G.-B. Kim, S. Baek, S.-B. Choi, D.-J. Hyung, S.-W. Lee, S.-H. Sohn, S.-J. Kwon, M. Jin, Y. J. Seol, W. B. Chae, K. J. Choi, B.-S. Park, H.-J. Yu, and J.-H. Mun. Elucidating the triplicated ancestral genome structure of radish based on chromosome- level comparison with the Brassica genomes. *Theoretical and Applied Genetics*, pages 1–16, Mar. 2016.

H. Kitashiba, F. Li, H. Hirakawa, T. Kawanabe, Z. Zou, Y. Hasegawa, K. Tonosaki, S. Shirasawa, A. Fukushima, S. Yokoi, Y. Takahata, T. Kakizaki, M. Ishida, S. Okamoto, K. Sakamoto, K. Shirasawa, S. Tabata, and T. Nishio. Draft Sequences of the Radish (*Raphanus sativus L.) Genome. DNA Research*, 21(5):481–490, Oct. 2014.

Q. Kong, X. Li, C. Xiang, H. Wang, J. Song, and H. Zhi. Genetic Diversity of Radish (*Raphanus sativus L.*) Germplasm Resources Revealed by AFLP and RAPD Markers. *Plant Molecular Biology Reporter*, 29(1):217–223, Mar. 2011.

H. Lischer, L. E. Bioinformatics, and 2011. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. academic.oup.com.

D. B. Lowry, S. Hoban, J. L. Kelley, K. E. Lotterhos, L. K. Reed, M. F. Antolin, and A. Storfer. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, 17(2): 142–152, Dec. 2016.

N. Lü, K. Yamane, and O. Ohnishi. Genetic diversity of cultivated and wild radish and phylogenetic relationships among *Raphanus* and *Brassica* species revealed by the analysis of trnK/matK sequence. *Breeding Science*, 58(1):15–22, 2008.

J. McNeill. The taxonomy and evolution of weeds. *Weed Research*, 16:399–413, Apr. 1976.

Y. Mitsui, M. Shimomura, K. Komatsu, N. Namiki, M. Shibata-Hatta, M. Imai, Y. Katayose, Y. Mukai, H. Kanamori, K. Kurita, T. Kagami, A. Wakatsuki, H. Ohyanagi, H. Ikawa, N. Minaka, K. Nakagawa, Y. Shiwa, and T. Sasaki. The radish genome and comprehensive gene expression profile of tuberous root formation and development. *Nature Publishing Group*,:1–14, June 2015.

G. D. Moghe, D. E. Hufnagel, H. Tang, Y. Xiao, I. Dworkin, C. D. Town, J. K. Conner, and S. H. Shiu. Consequences of Whole-Genome Triplication as Revealed by Comparative Genomic Analyses of the Wild Radish *Raphanus raphanistrum* and Three Other Brassicaceae Species. *The Plant Cell*, 26(5):1925–1937, May 2014.

H. Müller-Schärer, U. Schaffner, and T. Steinger. Evolution in invasive plants: implications for biological control. *Trends in Ecology & Evolution*, 19(8):417–422, Aug. 2004.

J. Muminovi ́c, A. Merz, A. E. Melchinger, and T. Lübberstedt. Genetic structure and diversity among radish varieties as inferred from AFLP and ISSR analyses. *Journal of the American Society for Horticultural Science*, 130(1):79–87, 2005.

P. Neve, M. Vila-Aiub, and F. Roux. Evolutionary-thinking in agricultural weed management. *New Phytologist*, 184(4):783–793, Sept. 2009.

R. Nielsen. Molecular signatures of natural selection. *Annual Review of Genetics*, 39: 197–218, 2005.

J. R. Paris, J. R. Stevens, and J. M. Catchen. Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*, 188:799–14, Apr. 2017.

N. Patterson, A. L. Price, and D. Reich. Population Structure and Eigenanalysis. *PLoS Genetics*, 2(12):2074–2093, 2006.

D. Pimentel, R. Zuniga, and D. Morrison. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, 52(3):273–288, Feb. 2005.

J. Pritchard and M. Stephens. Inference of population structure using multilocus genotype data. *Genetics*, 155:945–959, 2000.

J. K. Pritchard, X. Wen, and D. Falush. Documentation for STRUCTURE software: Version 2.3. University of Chicago, 2010.

R Core Team. R: A Language and Environment for Statistical Computing. 2015.

D. N. Reznick and C. K. Ghalambor. The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica*, 112-113:183–198, 2001.

H. F. Sahli, J. K. Conner, F. H. Shaw, S. Howe, and A. Lale. Adaptive Differentiation of Quantitative Traits in the Globally Distributed Weed, Wild Radish (*Raphanus raphanistrum*). *Genetics*, 180(2):945–955, Sept. 2008.

D. Shen, H. Sun, M. Huang, Y. Zheng, X. Li, and Z. Fei. RadishBase: a database for genomics and genetics of radish. *Plant and cell physiology*, 54(2):e3–e3, Feb. 2013a.

D. Shen, H. Sun, M. Huang, Y. Zheng, Y. Qiu, X. Li, and Z. Fei. Comprehensive analysis of expressed sequence tags from cultivated and wild radish (*Raphanus spp.*). *BMC Genomics*, 14(1):72, 2013b.

J. M. Smith and J. Haigh. The hitch-hiking effect of a favourable gene. *Genetical research*, 23(01):23–35, Feb. 1974.

A. A. Snow and L. G. Campbell. Can feral radishes become weeds? In Crop Ferality and Volunteerism, pages 193–208. *CRC Press*, Apr. 2005.

C. N. Stewart Jr, P. J. Tranel, D. P. Horvath, J. V. Anderson, L. H. Rieseberg, J. H. Westwood, C. A. Mallory-Smith, M. L. Zapiola, and K. M. Dlugosch. Evolution of Weediness and Invasiveness: Charting the Course for Weed Genomics. *Weed Science,* 57(5):451–462, Sept. 2009.

S. Sutherland. What makes a weed a weed: life history traits of native and exotic plants in the USA. *Oecologia,* 141(1):24–39, Aug. 2004.

P. J. Tranel and T. R. Wright. Resistance of weeds to ALS-inhibiting herbicides: what have we learned? *Weed Science*, 50(6):700–712, Nov. 2002.

C. C. Vigueira, K. M. Olsen, and A. L. Caicedo. The red queen in the corn: agricultural weeds as models of rapid adaptive evolution. *Heredity*, 110:303–311, Nov. 2012.

N. Wang, N. Kitamoto, R. Ohsawa, and T. Fujimura. Genetic diversity of radish (*Raphanus sativus*) germplasms and relationships among worldwide accessions analyzed with AFLP markers. *Breeding Science*, 58(2):107–112, 2008.

S. I. Warwick and A. Francis. The biology of Canadian weeds. 132. *Raphanus raphanistrum L. Canadian Journal of Plant Science*, 85(3):709–733, July 2005.

S. I. Warwick and C. N. Stewart. Crops come from wild plants: how domestication, transgenes, and linkage together shape ferality. In Crop Ferality and Volunteerism, pages 9–30. *CRC Press*, Apr. 2005.

T. D. Wu and C. K. Watanabe. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, 21(9):1859–1875, May 2005.

H. Yamagishi and T. Terachi. Multiple origins of cultivated radishes as evidenced by a comparison of the structural variations in mitochondrial DNA of *Raphanus*. *Genome,* 46(1):89–94, Feb. 2003.

K. Yamane, N. Lü, and O. Ohnishi. Chloroplast DNA variations of cultivated radish and its wild relatives. *Plant Science*, 168(3):627–634, Mar. 2005.

K. Yamane, N. Lü, and O. Ohnishi. Multiple origins and high genetic diversity of cultivated radish inferred from polymorphism in chloroplast simple sequence repeats. *Breeding Science*, 59:55–65, 2009.

**Chapter III: Differential Expression of floral traits under Artificial Selection**

## Introduction

Differences in floral morphology among angiosperms are often correlated with differences in important traits such as preferred pollinator (Bradshaw and Schemske, 2003; Harder and Barrett, 1993; Schemske, 2010), and degree of out-crossing (Barrett, 2003; Fornoni et al., 2015; Snell and Aarssen, 2005). As floral morphology directly impacts plant fitness, the genes for floral characters and development have been extensively studied in the model plant Arabidopsis thaliana (e.g., Abraham et al., 2013; Alvarez-Buylla et al., 2010; Grillo et al., 2013; Preston et al., 2011; Smyth et al., 1990; Yanofsky), and are relatively well studied even in non-model plants (as reviewed in: Endress, 2012; Glover et al., 2015; Kramer, 2007; Soltis et al., 2009).

Although the genetic mechanisms that underlie qualitiative shifts in floral traits are well understood, and the relative sizes of floral organs between plant species are frequently measured and reported in the literature as part of pollinator and development studies (e.g., Andersson, 1996; Bateman et al., 2013; Greyson and Sawhney, 2015; Humeau and Thompson, 2001; Niklas, 1994; Ushimaru and Nakata, 2015), little attention has been paid to determining the genetic basis of allometric changes, or changes in the relative sizes of floral structures between species. While there have been a small number of QTL studies where the relative sizes of floral traits have been assessed (Conner, 2002; Feng et al., 2009), we know of no studies that have identified the genes underlying these traits, although several genes underlying floral organ identity and overall size have been identified (Bowman et al., 1989; Hepworth and Lenhard, 2014; Krizek, 2009). We have attempted to fill this gap by finding candidate genes for allometric differences in a single species of Brassicaceae.

Brassicaceae contains more than 3700 plant species ranging from Arabidopsis thaliana, to cabbage (Brassica oleracea), to radish (*Raphanus raphanistrum*) to money plants (Lunaria annua). Although these species vary widely in terms of size, leaf structure, and several other traits, floral morphology is largely homogenous across the family (Conner, 2006; Franzke et al., 2011). Originally known as Cruciferae, the Brassicaceae typically have a "cross-like" corolla, consisting of 4 petals converging towards the base of the flower to form a tube. Each flower has 6 stamens; 4 tall and 2 short (Figure 3-1). Among Brassicaceae, absolute stamen and tube length vary by species. However most of this variation is accounted for by overall flower size, with most species having roughly the same stamen to tube ratio (Conner, 2006). This conservation across the taxon suggests the relative height of the stamens to the floral tube, a composite trait known as anther exsertion, may be adaptive. The anthers, which deposit pollen on insects, are borne on the tips of the stamens. As such, the degree of anther exsertion can have a profound
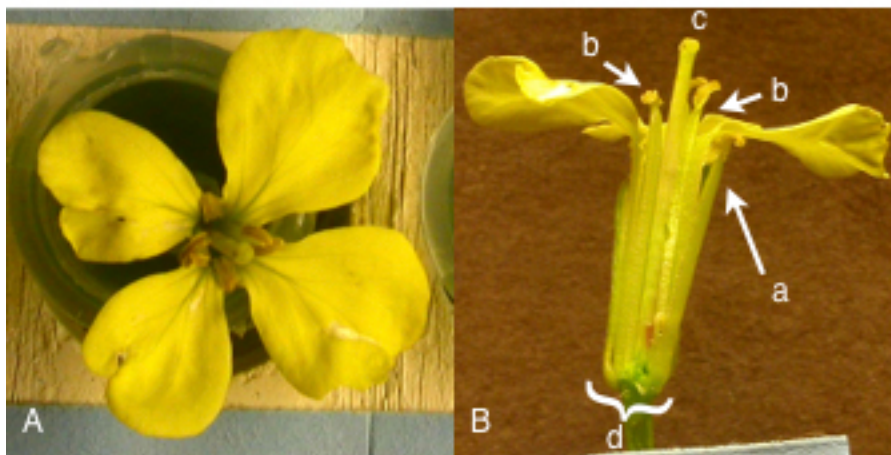


**Figure 3-1: Floral morphology of radish after partial dissection.** A. Top view of a radish flower showing cross-like petals, anthers of the 4 tall stamens, and the pistil. B. Side view of same flower after dissection showing a short stamen (a), 2 tall stamens (b), pistil (c), and the tube created by the petals (d).

impact on the efficiency and placement of pollen deposition on pollinators (Armbruster et al., 2014; Harder and Barrett, 1993; Kudo, 2003; Nishihiro and Washitani, 2011; Webb and Lloyd, 1986), which in turn can affect fitness (Conner, 2006).

In radish, *Raphanus raphanistrum ssp. raphanistrum*, pleiotropy causes a higher correlation between the height of the stamens and the height of the corolla tube than there are among other floral or vegetative traits (Conner, 2002; Conner and Sterling, 1995; Conner and Via, 1993). However, experiments that directly selected for increased or decreased height of the tall stamens relative to the corolla tube in radish had a large and rapid response, demonstrating that high genetic correlations don't necessarily result in strong evolutionary constraint (Conner, 2006; Conner et al., 2014). After eight and nine generations of selection in two replicate experiments, Conner et al. found differences of 3.4 and 5.8 standard deviations between the High and Low selection lines, respectively (Conner et al., 2011).

This system provides an excellent framework for finding genes that underlie quantitative allometric phenotypes. *R.r. raphanistrum* is an agricultural weed with the typical floral structure of Brassicaceae, and an emerging model for studying both rapid evolution and floral morphology (Agrawal, 1998; Agrawal et al., 2002, 2004; Bett and Lydiate, 2003; Campbell et al., 2006, 2009; Conner et al., 2011, 2014; Devlin and Ellstrand, 1990; Irwin and Strauss, 2005; Irwin et al., 2003; Klinger et al., 1991; Mazer and Schick, 1991; Morgan and Conner, 2001; Ridley and Ellstrand, 2009; Sahli et al., 2008; Stanton et al., 1986). As such, there are a number of genomic tools available for weedy radish and its close relatives. Both a transcriptome and a draft genome have been published for *R. r. raphanistrum* (Moghe et al., 2014), and crop radish, *R. sativus*, has

also been sequenced with a number of draft genomes available (Jeong et al., 2016; Kitashiba and Nasrallah, 2014; Mitsui et al., 2015).

To identify candidate loci controlling stamen and corolla tube height in radish, we did RNAseq on 35 plants from the artificial selection lines generated for (Conner et al., 2011, 2014) (17 low, 18 high), and performed a differential expression analysis. We mapped to all available reference genomes to provide a robust list of candidate genes underlying one floral trait, anther exsertion, in the weed *Raphanus raphanistrum*. We expect these results to be the first assessment of genes underlying allometric differences in floral traits, and to be broadly useful to the plant research community.

## Methods

### Experimental Setup

Plants from the Binghamton NY (BINY) population (42.184089E, 75.835319W), were selected as described in (Conner et al., 2011, 2014). Conner et al. used this population to create two replicate anther exsertion selection experiments. Each replicate contains two selection lines, one which was selected for a high stamen to corolla tube ratio – the "High" lines – and one with selection for a low ratio – the "Low" lines. In both the High and Low lines, selection was perpendicular to the correlation between filament and corolla tube lengths, such that the composite trait of anther exsertion was under directional selection. Each of these four selection lines was comprised of 12 maternal lines.

**RNA sequencing and mapping**

After six (Rep 1) or seven (Rep 2) generations of artificial selection, RNA was extracted from flower buds from 36 plants: nine from each of the lines. Although buds were not age-matched, all were early to mid-stage. Due to greenhouse space constraints, these plants were grown in blocks, approximately two months apart: all Rep 1 buds were collected in early February of 2008, while all Rep 2 buds were collected in late March of 2008. Buds were flash frozen in liquid nitrogen, and all RNA preps were done concurrently after the second bud collection.

RNA was sequenced using an Illumina Genome Analyzer II to obtain approximately 6.5 million reads for each of thirty-five individuals. Sequencing from one individual from Rep 1 Low failed and is not included in subsequent analysis. This is very early Illumina data, and as such has much shorter reads (36bps), lower depth, and somewhat higher error rates than current technologies. To account for this, we mapped our reads using GSNAP (Wu and Watanabe, 2005), which was initially designed for mapping mRNA reads. GSNAP uses soft-clipping, which allows it to dynamically ignore bases at the beginning or end of the read where most errors occur, and importantly, it can align reads as short as 14bp (Wu and Nacu, 2010). As such, 36 base pair reads can be mapped, even if they require extensive soft clipping. We used a kmer size of 15, allowed splicing, and kept only the single best match for each read.

To improve our ability to find differentially expressed genes, we mapped these reads to a variety of genomic resources. Two of these, the *R.r. raphanistrum* genome and transcriptome (Moghe et al., 2014), were created using the BINY population, and so should provide the best match for mapping. However, both of these resources are draft sequences, and neither covers the entire genome/transcriptome (Table 3-1). Therefore, we also mapped reads to three newer publicly

available *R. sativus* draft genomes (Jeong et al., 2016; Kitashiba et al., 2014; Mitsui et al., 2015)

which have higher genomic coverage (Table 3-1).

Mapped reads were quantified using the 'union' flag in HTseq v.0.6.1 (Anders et al., 2015) to

determine the raw number of reads mapped to each gene. All mapping and quantification was

performed per individual.

**Table 3-1: Genomic resources available for *Raphanus*.** Each genomic resource is a column and is referred to by its column name throughout this manuscript. Both *R.r. raphanistrum* resources were created using the same base population as this study: BINY. All three *R. sativus* genomes were created using daikon type varieties. Kitashiba et al. (2014) and Mitsui et al. (2015) sequenced the same daikon cultivar, however they used single haploid (S-h) and doubled haploid (D-h) stocks, respectively. To make these resources more easily comparable, the genome values reported here all use a minimum assembled sequence length of 100 base pairs rather than the summary statistics reported in each paper. These values are taken from Table S5 of Jeong et al. (2016).

| | **Moghe2014** | **RR3NYEST** | **Kitashiba2014** | **Mitsui2015** | **Jeong2016** |
|---|---|---|---|---|---|
| **Species** | *R.r. raphanistrum* pop: BINY | *R.r. raphanistrum* pop: BINY | *R. sativus cv. Aokubi S-h* | *R. sativus cv. Aokubi D-h* | *R. sativus cv. WK10039* |
| **Resource Type** | Genome | Transcriptome | Genome | Genome | Genome |
| **Number of Contigs** | 68331 | 10148 | 107492 | 72909 | 38732 |
| **N50 (kb)** | 10.1 | - | 10.7 | 7.1 | 19.6 |
| **Assembled Length (Mb)** | 254.6 | - | 402.3 | 383.3 | 426.2 |
| **Gene Models** | 38174 | 44290 ESTs | 80521 | 64657 | 46514 |
| **Reference** | Moghe et al. (2014) | Moghe et al. (2014) | Kitashiba et al. (2014) | Mitsui et al. (2015) | Jeong et al. (2016) |

**Modeling of differential expression**

We used DESeq2 (Love et al., 2018) to determine which genes show differential expression across various treatments. This software is suitable for working with low depth data, with its inherently increased uncertainty in expression estimates. We expect that most RNAseq loci are not differentially expressed, and the variance at loci with low expression levels will be higher, and thus expression more difficult to measure accurately, than at more highly expressed loci (Anders and Huber, 2010). DESeq2 deals with this by filtering out loci with too few reads and using a shrinkage analysis which uses information from all the samples for a given locus to reduce the variance at lowly expressed loci and give better estimates (Anders and Huber, 2010; Dillies et al., 2013; Seyednasrollah et al., 2013). This approach is particularly well suited to this dataset for two reasons. First, these samples were sequenced in 2008, and as such have much smaller overall read counts than a typical modern RNAseq experiment. As such, lowly expressed loci will have very low numbers of reads, and even loci that were of moderate expression in the cell will have comparatively low absolute read counts. Second, although this data set has relatively low read counts for any given individual, there are at least eight individuals from each of the four selection lines. As shrinkage analysis works better with more biological replicates to 'borrow' information from, the replication in this dataset should substantially offset the relatively low total read count.

To determine which genes were differentially expressed, we created and analyzed five datasets. Each dataset consisted of the concatenated raw read counts of all 35 individuals mapped to a single genome, and was analyzed using the model:

$$n_i \sim F + R + S \tag{1}$$

where *n* is the number of reads for individual *i*, *F* is a factor that accounts for combined variation due to flowcell and date of the sequencing runs, *R* is a factor representing the replicate, and *S* is a factor that describes the direction of artificial selection.

To determine which genes were consistently differentially expressed between the selection lines in both locations, we extracted the contrast of 'High vs. Low'. As we expect any changes in expression to be relatively small, and to maintain enough genes for downstream analysis, we used a lenient cutoff for differential expression. We kept all differentially expressed genes, regardless of fold change that had an adjusted p-value of <.1 as measured by a Wald test, and adjusted using the Benjamini-Hochberg (BH) adjustment (Benjamini and Hochberg, 1995) for subsequent analysis.

**Validation**

As each of the genomic resources are highly fragmented (Table 3-1), non-equivalent regions of the genome may not be represented in each. As our ability to detect differential expression is predicated on mapping, genome (or transcriptome) quality could affect what genes are detected. We compared our results from each genome at several stages: mapping, genes showing differential expression, GO term analysis, and collapsed GO term analysis.

To compare mapping rates, we used the total number of reads counted by HTseq per individual, divided by the number of raw reads in each sequencing file as the total reads per individual. As such, the denominator includes all unmapped, multi-mapped and low mapping quality ($< 20$) reads.

To compare genes showing differential expression, and to perform GO analyses, we converted all gene lists to a common set of gene identifiers. As only the Mitsui et al. (2015) authors published GO annotations for their genomic resource and was available in agriGO (see below) we converted the differentially expressed gene lists from all the others to the Mitsui et al. (2015) gene annotations by BLAST. We used the command line version of blastn (Nucleotide-Nucleotide BLAST 2.7.1+), with the default settings (evalue=10, wordsize >+4), and used the single best match as the homologous sequence. For all datasets except (Mitsui et al., 2015), this resulted in a small reduction of differentially expressed genes which did not have a BLAST match. We compared which genes were differentially expressed across the five datasets by taking the simple overlaps of the converted gene lists.

**GO Analysis**

We used agriGO (Tian et al., 2017) to perform a Singular Enrichment Analysis (SEA) on all of the differentially expressed genes for each of the five converted data sets. In each case, the list of differentially expressed genes was compared to a list of all possible genes. As overall mapping rates differed between genomes, we constrained the list of all possible genes for a given resource to contain only the genes that had at least one read mapped to them from that dataset. We compared which GO terms were over-represented among differentially expressed genes by taking the simple overlap of GO terms enriched for each dataset. As our differentially expressed genes were a small subset of the background, we used a Fisher's exact test to determine which GO terms were enriched, with an adjusted p-value (BH) of < 0.05. We set the minimum number of mapping entries to three, and used the complete GO ontology, based on the annotations from (Mitsui et al., 2015) already present in the agriGO database (access date 4/7/18.)

We used the webtools CateGOrizer (Hu et al., 2008) to simplify the GO terms for each dataset. In all cases, we input the GO terms that were enriched in the SEA (adjusted p-value (BH) of < 0.05), and collapsed the terms using the Plant GOslim classification and single count methods.

## Results

**Response to selection**

A detailed analysis of the response to selection can be found in Conner et al. (2011), however a few results are directly relevant here. In all four selection lines, anther exsertion responded in the desired direction, however in both reps, the High lines showed a stronger response than the Low lines as compared to the control for each replicate. In the High lines, the mean height of the long filament increased .44mm (4%) in Rep 1 and .32mm (3%) in Rep 2, while the mean heights of the corolla tube decreased by .56mm (6%) and 1.37mm (14%), respectively. In the Low lines, the mean height of the corolla tube increased by about 4% (.4mm) Rep 1 and increased by 6% (.62mm) in Rep 2. The filament in the Low lines decreased by .41mm (4%) and .8mm (8%), respectively.

Taken together, this resulted in an average difference in anther exsertion of $\Delta 1.79$mm, (3.4 standard deviations) in Rep 1. Rep 2 had a larger overall change with an average anther exsertion of 3.1mm, or a difference of about 5.8 standard deviations after eight or nine generations of selection (Fig. 3-2).

**Figure 3-2: Representative examples of selection for increased and decreased anther exsertion.** Flowers were partially dissected for photography. The middle flower is from the unselected base population. The lower left, and upper right show representatives of the Low and High line respectively

**Sequencing and Mapping**

Illumina sequencing resulted in four million to eight million reads per individual, and overall

mapping rates of these individuals varied widely between genomic resources (Fig. 3-3). In our

analysis, Jeong2016, the most recent genome, and RR3NYEST, the *R.r. raphanistrum*

transcriptome, performed poorly, with only about 20% of reads uniquely mapping for any

individual. Our reads mapped to the three older genomes at approximately the same rate (60%)

in all individuals. When compared in the common Mitsui2015 background, the majority of gene

models in each background had at least one mapped read and overlap between mapped reads is

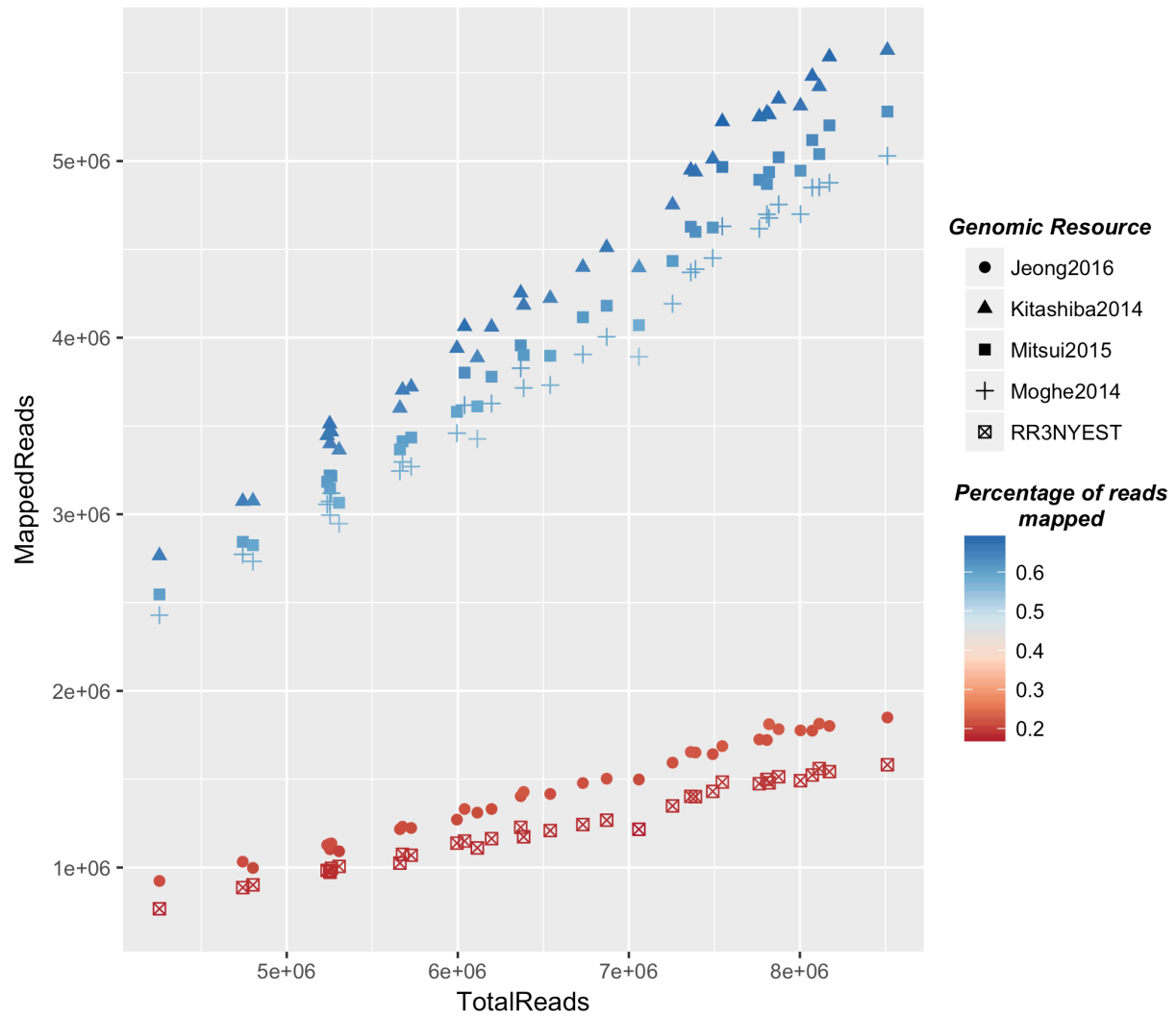relatively high across genomic resources (Fig. 3-4).

**Figure 3-3: Comparison of mapping rates across genomes.** For each of 35 individuals, the number of reads that passed all mapping filters plotted against the total number of raw reads. Mapping percentage is indicated by color. Genomic Resources are defined in Table 3-1.
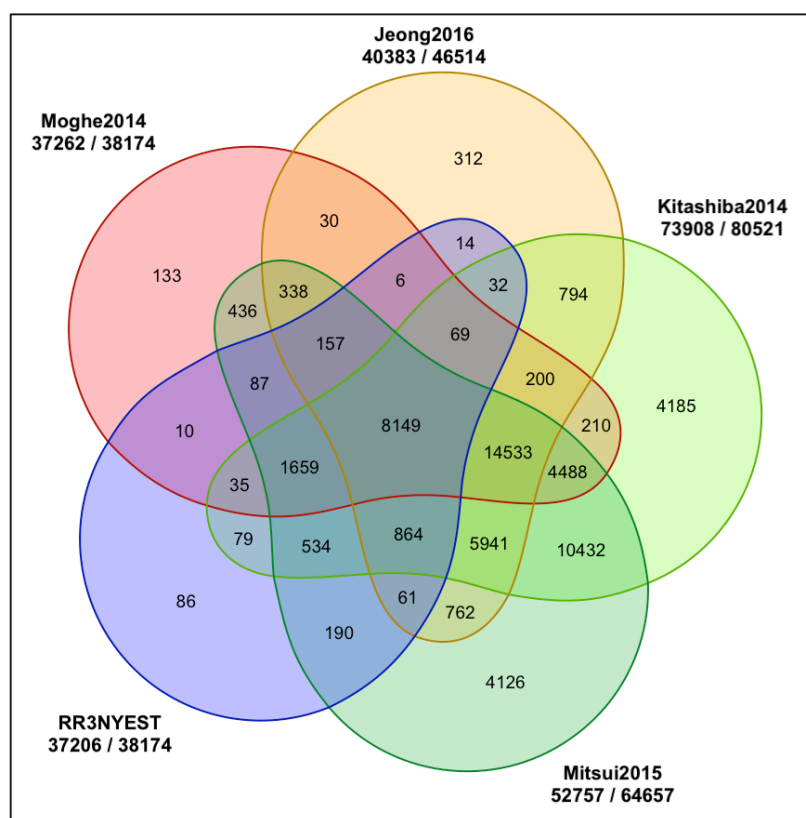
**Figure 3-4: Comparison of overall mapping to gene models by genomic resource.** This venn diagram shows the overlap for overall mapping among the genomic resources and includes all genes with at least one read mapped to them for a given genomic resource. As each resource uses an idiosyncratic gene model naming convention, all gene models for a given genomic resource (except Mitsui2005) were converted via blast to the annotations of Mitsui2015. Under each genomic resource is the total number of genes in that lobe of the venn followed by the number of gene models in that genomic resource. While overall mapping rates varied widely per genomic resource, the percentage of gene models with mapped reads was high across all resources.

**Differential Expression and GO analysis**

Between 3.1% and 5.9% of genes showed differential expression in each dataset, across genomic resources (Table 3-2). As the Low lines experienced relatively little phenotypic change, we are expressing the direction of change for each gene relative to the Low line. As such, we find that between 1.4% and 2.9% of genes are up-regulated in the High lines, and between 1.7% and 3% are down-regulated in the High lines. We observed only a small skew towards down regulation

**Table 3-2: Differentially expressed genes by genomic resource.** Differentially expressed genes with a Benjamini-Hochberg adjusted p-value of less than .1 and any fold change were kept for downstream analysis. The percentage of genes that experienced up (more reads in High lines) and down (more reads in Low lines) regulation were approximately the same regardless of genomic resource.

| Resource | Number Up Regulated (%) | Up, mean log2FC (SD) | Number Down Regulated (%) | Down, mean log2FC (SD) |
|---|---|---|---|---|
| Moghe2014 | 1164 (2.9) | 0.80 (0.76) | 1219 (3) | -0.83 (0.75) |
| RR3NYEST | 878 (2.2) | 1.07 (0.73) | 989 (2.4) | -1.12 (0.80) |
| Kitashiba2014 | 1517 (2) | 1.06 (0.84) | 1800 (2.4) | -1.08 (0.95) |
| Mitsui2015 | 1231 (2.3) | 1.0 (0.85) | 1509 (2.9) | -0.98 (0.91) |
| Jeong2016 | 605 (1.4) | 0.98 (0.85) | 712 (1.7) | -1.02 (0.96) |

**Table 3-3: The twenty most significantly enriched GO terms for up-regulated genes shared among all genomic resources.** Only GO terms from the Cellular Process (P) and Cellular Component (C) ontologies were shared across all data sets, values for each genomic resource are BH adjusted p-values.

| GO Term | On | Description | Moghe2014 | RR3NYest | Kitashiba2014 | Mitsui2015 | Jeong2016 |
|---|---|---|---|---|---|---|---|
| GO:0005746 | C | mitochondrial respiratory chain | 2.50E-11 | 9.70E-07 | 7.40E-13 | 3.40E-17 | 4.10E-08 |
| GO:0070469 | C | respiratory chain | 8.00E-12 | 3.50E-06 | 2.80E-13 | 3.40E-17 | 4.10E-08 |
| GO:0044455 | C | mitochondrial membrane part | 5.30E-13 | 2.50E-08 | 5.00E-14 | 1.20E-17 | 6.10E-06 |
| GO:0080129 | P | proteasome core complex assembly | 1.90E-07 | 6.80E-05 | 1.40E-11 | 2.60E-14 | 6.30E-07 |
| GO:0005740 | C | mitochondrial envelope | 2.50E-10 | 4.20E-08 | 5.50E-13 | 6.10E-15 | 0.00012 |
| GO:0044429 | C | mitochondrial part | 1.70E-09 | 4.20E-08 | 2.90E-11 | 3.90E-13 | 0.00012 |
| GO:0005743 | C | mitochondrial inner membrane | 5.50E-10 | 5.20E-08 | 5.60E-12 | 1.00E-14 | 0.00012 |
| GO:0031966 | C | mitochondrial membrane | 2.50E-09 | 1.90E-07 | 6.80E-11 | 1.40E-14 | 0.00012 |
| GO:0010038 | P | response to metal ion | 1.40E-05 | 8.60E-05 | 2.10E-08 | 1.50E-08 | 3.80E-05 |
| GO:0051788 | P | response to misfolded protein | 2.60E-06 | 0.00017 | 7.80E-10 | 7.70E-13 | 1.20E-06 |
| GO:0031090 | C | organelle membrane | 8.40E-06 | 7.70E-09 | 1.90E-09 | 8.10E-09 | 0.00029 |

Table 3-3 (cont'd)

| GO:0044425 | C | membrane part | 0.00013 | 5.10E-05 | 1.80E-09 | 4.80E-08 | 0.00012 |
|---|---|---|---|---|---|---|---|
| GO:0005747 | C | mitochondrial respiratory chain complex I | 7.80E-09 | 0.00033 | 3.60E-11 | 1.30E-13 | 6.10E-06 |
| GO:0006066 | P | alcohol metabolic process | 6.10E-08 | 0.00033 | 6.90E-11 | 4.80E-10 | 2.60E-05 |
| GO:0046686 | P | response to cadmium ion | 9.10E-06 | 0.00033 | 5.00E-09 | 4.60E-08 | 2.60E-05 |
| GO:0006096 | P | glycolysis | 0.00011 | 7.80E-05 | 2.60E-09 | 1.40E-08 | 0.0002 |
| GO:0043248 | P | proteasome assembly | 4.20E-06 | 0.00046 | 1.20E-09 | 1.20E-12 | 1.20E-06 |
| GO:0045271 | C | respiratory chain complex I | 2.50E-09 | 0.00048 | 1.40E-11 | 3.10E-14 | 6.10E-06 |
| GO:0030964 | C | NADH dehydrogenase complex | 2.50E-09 | 0.00048 | 1.40E-11 | 3.10E-14 | 6.10E-06 |
| GO:0005774 | C | vacuolar membrane | 0.00039 | 2.80E-09 | 1.20E-08 | 2.60E-07 | 0.00012 |

**Table 3-4: The twenty most significantly enriched GO terms for down-regulated genes shared among all genomic resources.** Only GO terms from the Cellular Process (P) and Cellular Component (C) ontologies were shared across all data sets, values for each genomic resource are BH adjusted p-values.

| GO Term | On | Description | Moghe2014 | RR3NYest | Kitashiba2014 | Mitsui2015 | Jeong2016 |
|---|---|---|---|---|---|---|---|
| GO:0044435 | C | plastid part | 1.40E-08 | 7.60E-14 | 2.80E-16 | 1.40E-08 | 4.30E-07 |
| GO:0044434 | C | chloroplast part | 1.40E-08 | 3.70E-13 | 4.30E-16 | 1.60E-08 | 4.30E-07 |
| GO:0009507 | C | chloroplast | 1.40E-08 | 2.60E-14 | 3.20E-13 | 3.00E-07 | 3.60E-07 |
| GO:0009536 | C | plastid | 5.40E-08 | 1.30E-13 | 2.30E-12 | 1.00E-06 | 3.60E-07 |
| GO:0009532 | C | plastid stroma | 4.20E-07 | 7.60E-14 | 7.00E-14 | 1.00E-06 | 2.70E-06 |
| GO:0009570 | C | chloroplast stroma | 4.20E-07 | 3.70E-12 | 1.00E-12 | 1.30E-06 | 4.40E-06 |
| GO:0009941 | C | chloroplast envelope | 3.80E-05 | 4.80E-07 | 9.20E-08 | 9.30E-06 | 1.10E-06 |
| GO:0009526 | C | plastid envelope | 4.20E-05 | 8.70E-07 | 1.40E-07 | 6.80E-06 | 4.30E-07 |
| GO:0031967 | C | organelle envelope | 0.00047 | 2.50E-05 | 6.40E-06 | 5.90E-08 | 5.20E-06 |
| GO:0006091 | P | generation of precursor metabolites and energy | 0.00052 | 1.70E-06 | 3.10E-07 | 3.10E-06 | 3.40E-05 |
| GO:0034660 | P | ncRNA metabolic process | 0.00035 | 1.60E-12 | 5.00E-15 | 0.00026 | 8.30E-07 |
| GO:0031975 | C | envelope | 0.00059 | 3.60E-05 | 9.30E-06 | 5.90E-08 | 5.20E-06 |

Table 3-4 (cont'd)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| GO:0016072 | P | rRNA metabolic process | 0.00049 | 1.00E-09 | 1.40E-10 | 0.00039 | 4.60E-06 |
| GO:0006364 | P | rRNA processing | 0.00077 | 8.60E-10 | 1.10E-10 | 0.00032 | 3.80E-06 |
| GO:0042254 | P | ribosome biogenesis | 0.001 | 1.50E-09 | 1.20E-09 | 0.00023 | 3.10E-05 |
| GO:0022613 | P | ribonucleoprotein complex biogenesis | 0.0015 | 1.50E-09 | 1.20E-09 | 0.00025 | 2.70E-05 |
| GO:0009658 | P | chloroplast organization | 0.0011 | 2.50E-09 | 1.60E-12 | 0.0004 | 0.00064 |
| GO:0055086 | P | nucleobase, nucleoside and nucleotide metabolic process | 0.00077 | 1.20E-07 | 1.30E-08 | 0.0012 | 0.00024 |
| GO:0034470 | P | ncRNA processing | 0.00086 | 8.30E-11 | 3.30E-12 | 0.0017 | 5.20E-06 |
| GO:0031984 | C | organelle subcompartment | 3.40E-06 | 9.20E-11 | 2.70E-12 | 0.0035 | 5.20E-06 |

in terms of gene number (Table 3-2) but no obvious differences in magnitude of change between up and down regulated genes (Table 3-2, Fig 3-5).

While the normalized gene counts in differentially expressed genes are consistently different between selection lines, we find a number of expression patterns. For some genes, we see a large overall expression difference between replicates, coupled with a smaller difference between selection lines (Fig. 3-6, A). In several cases, there is a clear pattern of differential expression between replicates, however the effect is stronger in one replicate, usually Rep 2 (Fig. 3.6, B-D). Finally, we also find many genes with the expected trend of a large difference in normalized expression shared across both replicates (Fig. 3-6, E & F). Note that in some of these cases, the trend of differential expression can also be characterized as a change in variability among samples within a selection line, as in figure 3-6 B, where the Low samples from Rep 1 are much less variable than Rep 2.

Although there was relatively little overlap between genes showing differential expression in each genome (Fig. 3-7, left), overlap between GO terms based on these genes was relatively high (Fig. 3-7, right). Between the five datasets, 472 GO terms are enriched in up-regulated genes (BH adjusted p value < .05), however about half of these (239) were enriched in only one or two genomic resources (Fig. 3-8, left). Similarly, we found 770 GO terms enriched among down-regulated genes (BH adjusted p value < .05), only 370 of which were shared among at least three genomic resources (Fig. 3-8, right). GO terms showing the most significant enrichment among up-regulated genes showed a clear bias towards mitochondrial annotations (Table 3-3), whereas down-regulated genes were overwhelmingly related to plastids (Table 3-4).
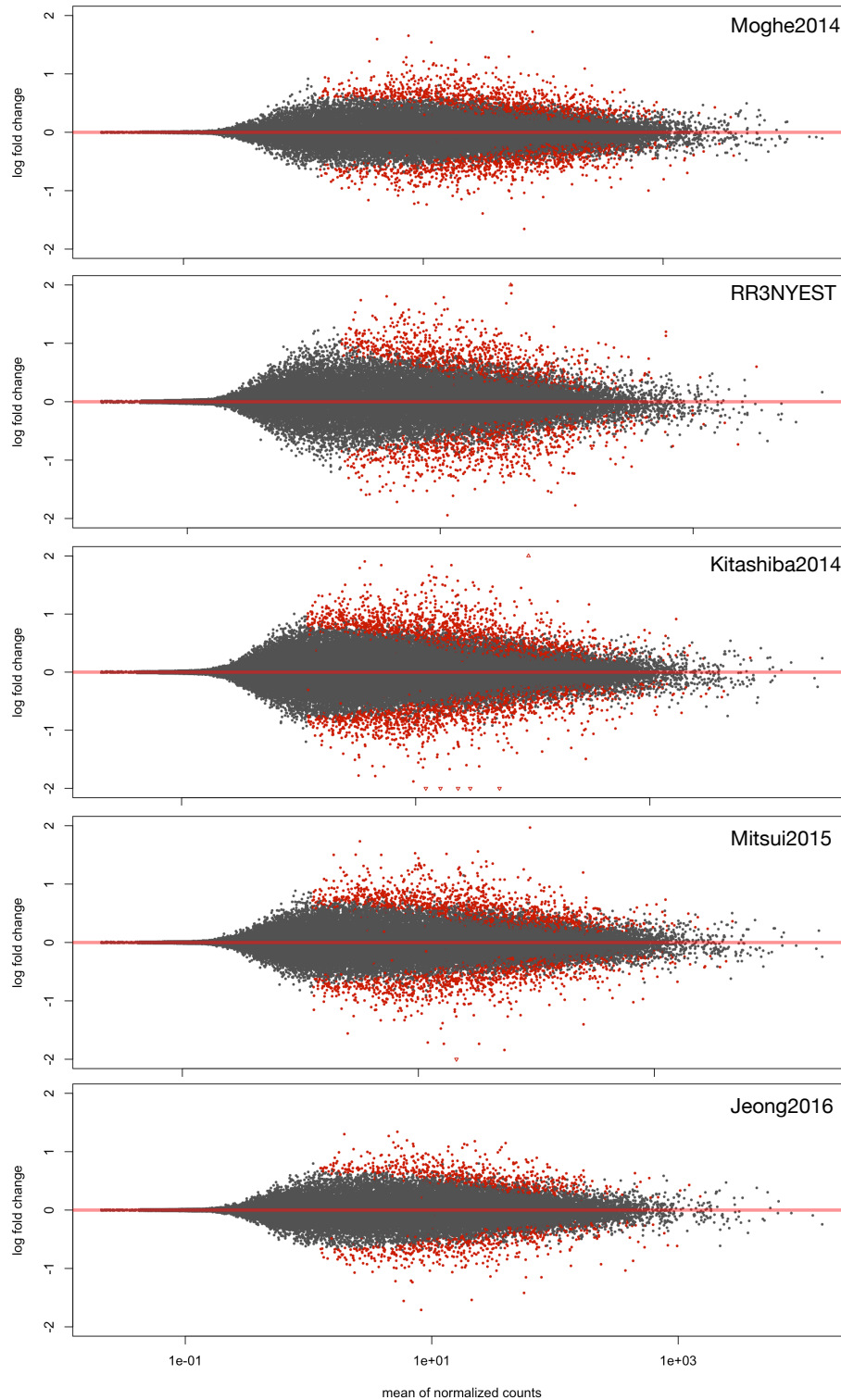
**Figure 3-5: Comparison of MA plots by genomic resource.** Plots show the distribution of differentially expressed genes by modeled read count and log fold change for each genomic resource: Moghe2014, RR3NYEST, Kitashiba2014, Mitsui2015 and (Top to bottom). Points in red are significant at p-adjusted (BH) p <.1
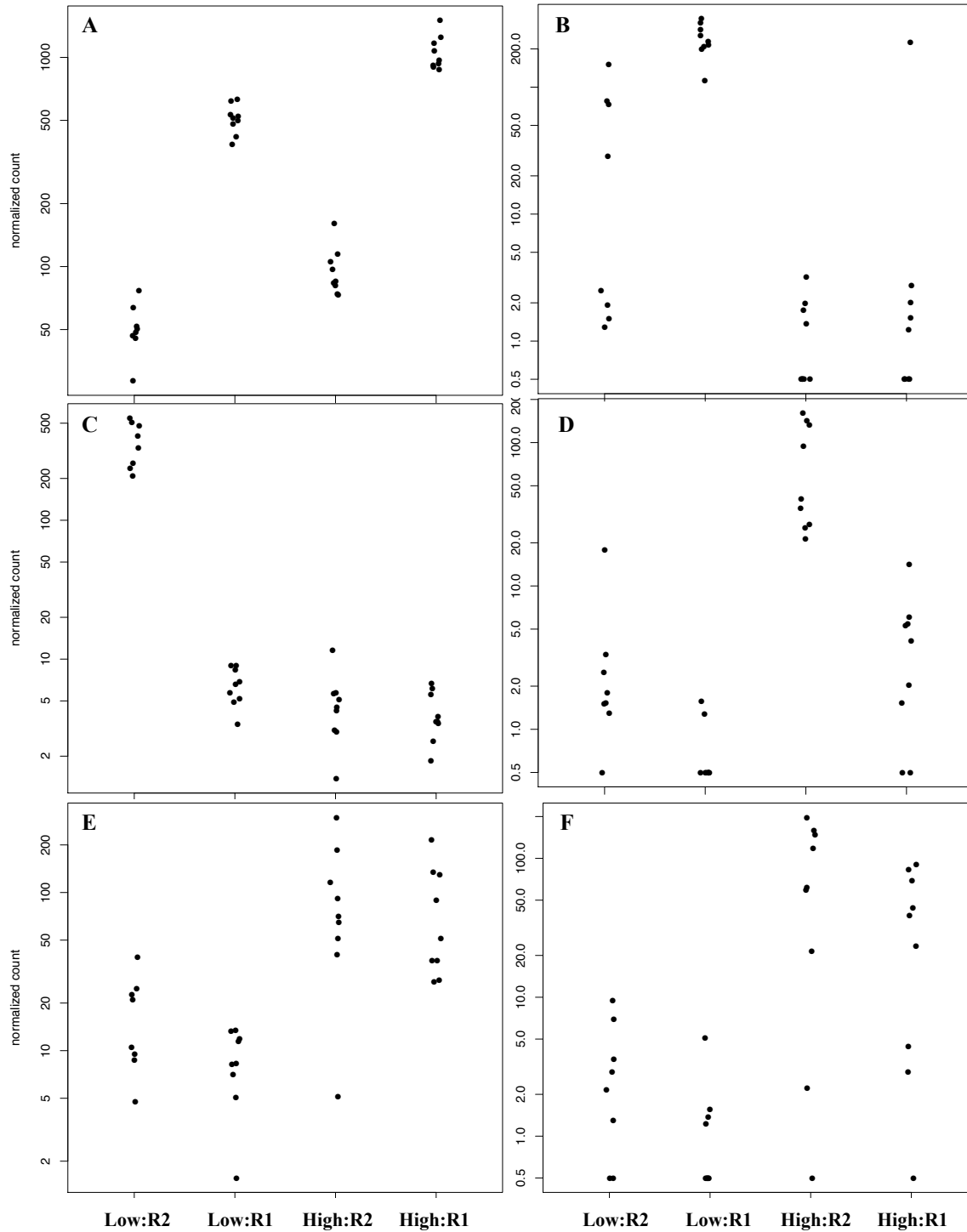
**Figure 3-6: Examples of normalized count values for six genes.** Each panel shows normalized counts for one of the six most significantly differentially expressed genes in the High and Low lines when mapped to Mitsui2015. Note that while the y-axis is normalized count values throughout, the scales are very different between plots. Each point is the normalized count for an individual.
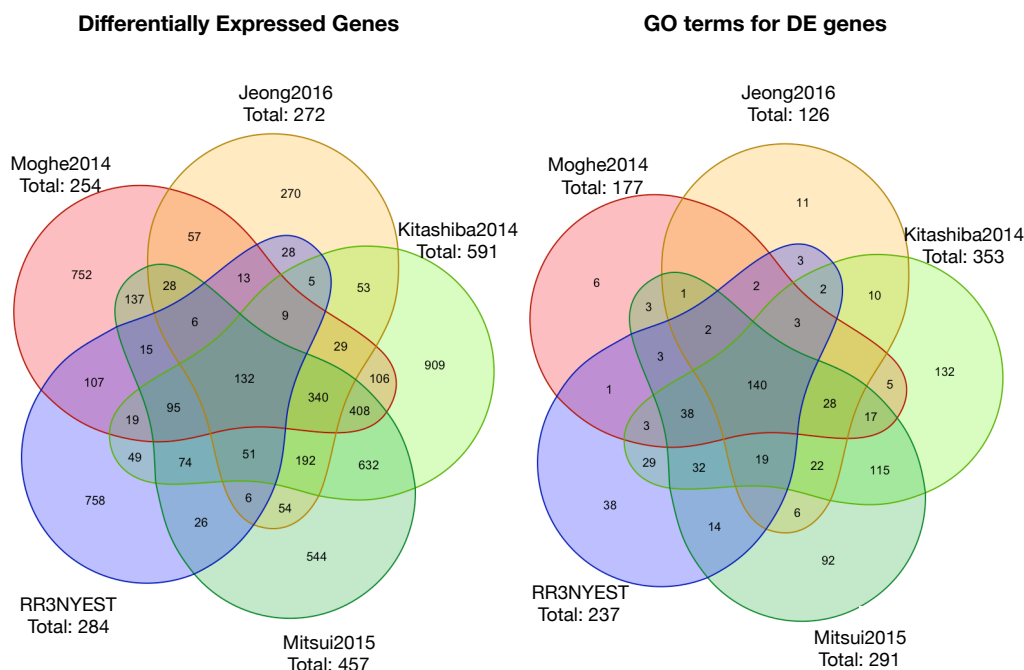
**Figure 3-7: Comparison of genes found to be differentially expressed by genomic resource, categorized by gene annotation and GO terms.** As each resource uses an idiosyncratic gene model naming convention, all gene models for a given genomic resource (except Mitsui2005) were converted via blast to the annotations of Mitsui2015. (Left) This shows the number of genes showing differential expression for the contrast of High vs Low from each genome. Under each genomic resource is the total number of genes in that lobe of the venn. For all resources except Mitsui2015, this number is somewhat less than the total of differentially expressed genes in Table 3-2, as some genes did not have blast hits in Mitsui2015. (Right) Overlap of GO annotations attributed to the differentially expressed genes above.
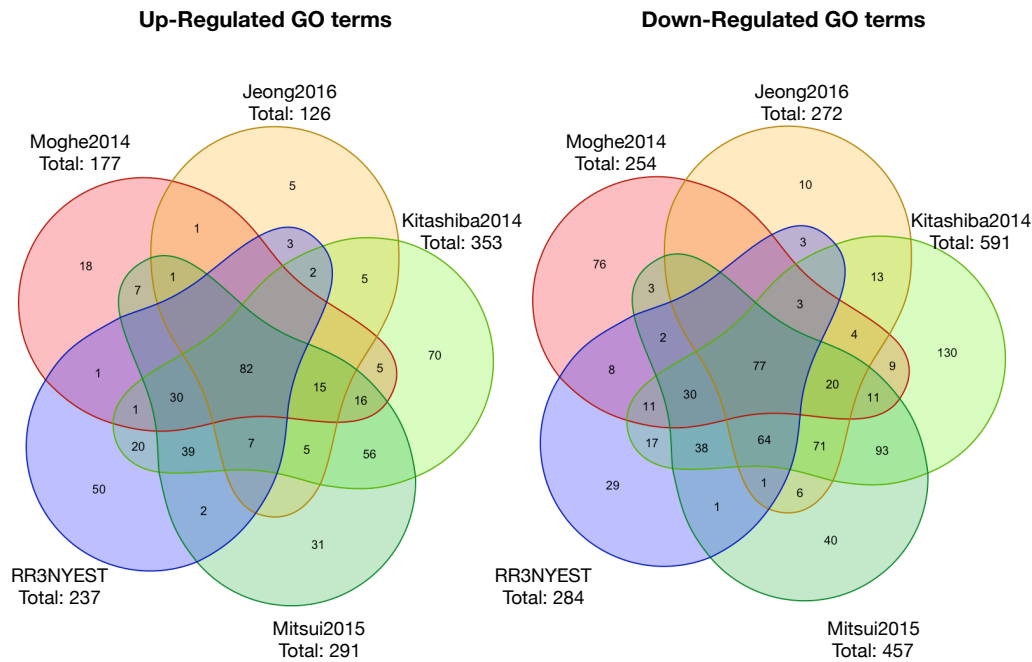
**Figure 3-8: Comparison of overlap among GO annotations for genes found to be up and down regulated by genomic resource.** These venn diagrams show the overlap for complete gene ontology annotations among the genomic resources that were enriched with a BH adjusted p-value of <.05. Under each genomic resource is the total number of GO annotations in that lobe of the venn.

## Discussion

We used RNAseq data to test for differential expression between buds from *R.r. raphanistrum* plants that had been artificially selected for High or Low anther exsertion, a composite trait directly related to pollination. Although there are a number of genomic resources available for *R. raphanistrum*, all are in draft form, and are highly fragmented. To improve the robustness of our experiment, we did five differential expression analyses in parallel. We mapped to both the *R.r. raphanistrum* draft genome (Moghe2014) and transcriptome (RR3NYEST) (Moghe et al., 2014),

and to three draft *R. sativus* genomes: Kitashiba:2014 (Kitashiba and Nasrallah, 2014),

Mitsui2015 (Mitsui et al., 2015), and Jeong2016 (Jeong et al., 2016).

**Mapping varied unexpectedly among resources**

For three of our genomic resources Moghe2014, Kitashiba2014 and Mitsui2014, we obtained

unique mapping rates of approximately 60% (Fig. 3-3). This is somewhat lower than is common

for a modern mapping experiment, however as we had very short reads and *Raphanus* has a

highly duplicated genome (Jeong et al., 2016; Moghe et al., 2014), in many thousands of

fragments (Table 3-1), this is not unexpected.

However, only about 20% of our reads mapped to either Jeong2016 or RR3NYEST (Fig. 3-3).

This is surprising, as the Jeong2016 genome appears to be the most well assembled of our

resources (Table 3-1) and was reported to cover in excess of 90% of the gene space (Jeong et al.,

2016). Similarly, the RR3NYEST was built from the same population as the selection lines used

in this study, and so should have provided an excellent match.

Although mapping rates per genomic resource varied widely, individual mapping rates were

consistent within each genomic resource (Fig. 3-3). Similarly, although mapping rates and the

absolute number of gene models vary widely between genomic resources (Table 3-1), the

percentage of gene models with at least one read mapped to them were quite high and ranged

from 81-97% (Fig. 3-4). Kitashiba2014 and Mitsui2005, have a relatively large number of

unique gene models with mapped reads (Fig. 3-4), however these two resources also have

approximately twice as many proposed gene models as the other resources (Table 3-1). Thus, it

appears that mapping rates had little effect on our mapping coverage, and the low mapping rates

in Jeong2016 and RR3NYEST must have resulted in lower depths. We therefore suspect that in building the RR3NYEST and Jeong2016 resources, either true homologs were discarded, or some were inappropriately collapsed into single gene models. As our reads were very short, we used the default combined penalty score, which would allow a maximum of 2 mismatches, and so reads would not map well to moderately diverged homologous sequences.

**GO Analysis is similar across genomic resources**

With lenient cutoffs for significance (LFC > 0, BH adjusted p-value 0.1), we found that fewer than six percent of genes showed evidence of differential expression, regardless of the genomic resource used for mapping, with only a small bias towards down-regulation (Table 3-2). Plots of the normalized gene counts for these differentially expressed genes show a number of expression patterns. While differential expression in some genes appears to be dominated by only one replicate (Fig 3-6, B-D), others show a clear pattern of repeated expression changes across replicates (Fig 3-6, A, E &F).

When compared in the common Mitsui2015 background, we found relatively little overlap between genes showing differential expression (Fig. 3-7. left). However, converting these gene lists to GO terms substantially improved their overlap (Fig. 3-7. right). This is expected for several possible reasons; firstly, if we have detected biologically meaningful differential expression, we would expect to find genes involved in building floral organs or changing their relative size. In both cases, we would expect that these genes work as a integrated unit (Fornoni et al., 2015), and are likely to have similar GO terms.

We also suspect that some of this collapsing effect is due to the triplicated nature of the *Raphanus* genome. Duplicated genes in *R.r. raphanistrum* often undergo rapid divergence in sequence and expression (Moghe et al., 2014), and we expect that our mapping parameters were stringent enough to discern these paralogs. However, these changes do not necessarily imply changes in gene function (Nowak et al., 1997), and we expect that at least in some cases both genes will retain their original function. As GO terms are often created using genomically simpler species like *Arabidopsis*, there is also the possibility that some of these genes may actually have undergone neofunctionalization that was missed when assigning GO terms by BLAST.

**Understanding allometric effects**

The most significantly enriched GO terms across all genomic resources for genes up-regulated in the High lines show a strong bias towards the mitochondria and respiration (Table 3-3), whereas down-regulated genes are largely related to plastids, particularly the chloroplast (Table 3-4). In both groups, the enriched GO terms additionally suggest that the differences are in number of organelles rather than function. For example, seven of the top twenty up-regulated GO terms are specifically named as mitochondrial membrane components; and others, like the NADH dehydrogenase complex, are mitochondian membrane-bound. Together, this implies an overall increase in the number of mitochondria being manufactured (Table 3-3). Similarly, eight of the top twenty down-regulated GO terms suggest an overall reduction in the number of plastids (Table 3-4).

These overall results suggest we may be seeing signal from changes in cell size, as both number of mitochondria and number of chloroplasts have been shown to scale with cell volume (Okie et

al., 2016; Savage et al., 2007). According to metabolic allometric scaling rules (Miettinen and Björklund, 2017), larger cells have both more mitochondria and a lower metabolic rate; whereas smaller cells have fewer organelles, but a much higher metabolic rate (Aryaman et al., 2017; Savage et al., 2007).

Our sequencing data suggests an increased number of mitochondria coupled with an overall decrease in energy storage, which may indicate cells which are near starvation. While the reduction in plastids might suggest an overall reduction in cell size, which would increase energy needs, that would not explain the apparent increase in mitochondria. However, there are a few reasons to suggest that we may be seeing the effects of both smaller and larger cells simultaneously. First, we have confounded the effects of differential expression in these lines by not sequencing a set of controls. Although the High lines had a larger absolute change in anther exsertion, there was still a substantial reduction of relative anther height in the Low lines (Fig. 3-2). However, in our analysis, over-expression of a gene in the High lines is indistinguishable from a reduction in Low line expression. Second, the anther exsertion phenotype in the High lines has two components: the length of the tube decreased by 1.4mm, while the height of the stamens increased by .75mm. As we collected and extracted RNA from whole buds, any differential expression signal from the High lines will include genes for both reducing petal size and increasing filament length. RNAseq on the tube and filament organs separately, and the use of control lines, would help to deconvolute these signals.

Although we found a number of genes experiencing differential selection, we found almost no evidence for directional selection in genes that have been previously shown to control size and shape in plants (Bowman et al., 1989; Hepworth and Lenhard, 2014; Krizek, 2009; Powell and

Lenhard, 2012). Of the 2740 genes differentially expressed in our Mitsui2015 mapping, only thirty-one have annotations similar to these known organ size determinate genes, and only twenty have an adjusted p-value of $< .05$; none were in the top 250 differentially expressed genes. However, those studies primarily used mutant lines, or assayed mostly leaves, and none looked at the allometric changes in specific floral organs. As such, they may have seen more exaggerated or organ specific effects. However, as the shrinkage analysis in DeSeq2 reduces the variance of log2fold change, and our data was relatively low depth, we are also likely missing many differentially expressed genes.

We assayed early buds of *R.r. raphanistrum* for changes in gene expression correlated to artificial selection for increased and decreased anther exsertion. Studies from *Arabidopsis*, the closest relative for which we have data, suggest that our buds should have just made primordial stamens and petals (Smyth et al., 1990), and so we expected to capture differential expression in the genes that create these structures. However, other stages of floral development might be better suited for capturing changes in anther exsertion. In *Arabidopsis*, mid-stage buds have relatively long stamens, and the petals only reach the height of the long stamens as the bud opens. After the bud opens, the long stamens resume growth to exsert past the opening of the tube (Smyth et al., 1990). Late, up-opened buds, or very early flowers may show more signal in genes commonly associated with allometry, and sequencing across all these stages may be necessary to find all of the genes related to differences in relative anther length.

Further, although the relative degree of anther exsertion in Brassicaceae is stable across the family, and presumed to be highly constrained, the ratio was readily perturbed with only a few generations of artificial selection of *R.r. raphanistrum*. However, F2 offspring made by crossing

the High and Low lines showed a continuous distribution of traits (Conner et al., 2014). This suggests that selection for anther exsertion favored many small changes across the genome, which might be due to differential regulation of genes controlling loci for traits like cell size, shape or growth. A much larger experiment, with tissue from more floral growth stages would allow us to find more loci of small effect, as well as those that only contribute to floral morphology for short periods of development. Similarly, we sequenced buds after six or seven generations of artificial selection, whereas Conner et al. (2011) assayed differences in anther exsertion after eight or nine generations. Although the mean response to selection did not change substantially between six and nine (Conner et al., 2011), we may find stronger differential expression by re-sequencing plants from later generations.

In only a few generations of selection, we saw large differences in floral morphology with responses in both directions – despite the pre-existing genetic correlation. Although our study suffers from a lack of power due to multiple confounding issues, we find robust differences in expression between *R.r. raphanistrum* lines selected for both high and low anther exsertion that are likely a reliable measure of differentiation for anther exsertion in early buds. Indeed, even in cases where the effects of replicate are very large (Fig. 3-6, A), we were still able to detect consistent expression differences between selection lines.

High anther exsertion is correlated with both an increase in the expression of mitochondrial components and a decrease in expression of plastids. These expression differences suggest that selection in these lines favored a change in cell size of the floral tube and filament. It is perhaps unsurprising that we see a signal primarily in organelle production, as they are relatively large cellular components; and that the initial response to selection seems to be due to cell traits that

are able to vary without disruption to more complex/integrated genetic networks, hence the adaptation in cell size/organelle number which already vary among tissues and over developmental time.

**REFERENCES**

# REFERENCES

M. C. Abraham, C. Metheetrairut, and V. F. Irish. Natural variation identifies multiple loci controlling petal shape and size in Arabidopsis thaliana. *PLoS One*, 8(2):e56743, 2013.

A. A. Agrawal. Induced responses to herbivory and increased plant performance. *Science*, 279 (5354):1201–1202, Feb. 1998.

A. A. Agrawal, J. K. Conner, M. T. J. Johnson, and R. Wallsgrove. Ecological genetics of an induced plant defense against herbivores: additive genetic variance and costs of phenotypic plasticity. *Evolution*, 56(11):2206–2213, Nov. 2002.

A. A. Agrawal, J. K. Conner, and J. R. Stinchcombe. Evolution of plant resistance and tolerance to frost damage. *Ecology Letters*, 7(12):1199–1208, Dec. 2004.

E. R. Alvarez-Buylla, M. Benítez, A. Corvera-Poiré, Á. Chaos Cador, S. de Folter, A. Gamboa de Buen, A. Garay-Arroyo, B. García-Ponce, F. Jaimes-Miranda, R. V. Pérez-Ruiz, A. Piñeyro- Nelson, and Y. E. Sánchez-Corrales. Flower Development. *The Arabidopsis Book*, 8:e0127–57, Jan. 2010.

S. Anders and W. Huber. Differential expression analysis for sequence count data. *Genome Biology*, 11(10):R106, Oct. 2010.

S. Anders, P. T. Pyl, and W. Huber. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–169, Jan. 2015.

S. Andersson. Flower-fruit size allometry at three taxonomic levels in Crepis (Asteraceae). *Biological Journal of the Linnean Society*, 58(4):401–407, Aug. 1996.

W. S. Armbruster, S. A. Corbet, A. J. M. Vey, S.-J. Liu, and S.-Q. Huang. In the right place at the right time: Parnassia resolves the herkogamy dilemma by accurate repositioning of stamens and stigmas. *Annals of Botany*, 113(1):97–103, Jan. 2014.

J. Aryaman, H. Hoitzing, J. P. Burgstaller, I. G. Johnston, and N. S. Jones. Mitochondrial heterogeneity, metabolic scaling and cell death. *BioEssays*, 39(7):1700001–9, June 2017.

S. C. H. Barrett. Mating strategies in flowering plants: the outcrossing-selfing paradigm and beyond. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 358 (1434):991–1004, June 2003.

R. M. Bateman, K. E. James, and P. J. Rudall. Contrast in levels of morphological versus molecular divergence between closely related Eurasian species of Platanthera(Orchidaceae) suggests recent evolution with a strong allometric component. *New Journal of Botany*, 2(2):110–148, Nov. 2013.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, 57(1):289–300, 1995.

K. E. Bett and D. J. Lydiate. Genetic analysis and genome mapping in *Raphanus*. *Genome*, 46 (3):423–430, June 2003.

J. L. Bowman, D. R. Smyth, and E. M. Meyerowitz. Genes Directing Flower Development in Arabidopsis. *The Plant Cell*, 1(1):37–52, 1989.

H. D. Bradshaw and D. W. Schemske. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, 426(6963):176–178, Nov. 2003.

L. G. Campbell, A. A. Snow, and C. E. Ridley. Weed evolution after crop gene introgression: greater survival and fecundity of hybrids in a new environment. *Ecology Letters*, 9(11):1198–1209, Nov. 2006.

L. G. Campbell, A. A. Snow, P. M. Sweeney, and J. M. Ketner. Rapid evolution in crop-weed hybrids under artificial selection for divergent life histories. *Evolutionary Applications*, 2(2): 172–186, May 2009.

J. K. Conner. Genetic mechanisms of floral trait correlations in a natural population. *Nature*, 420 (6914):407–410, Nov. 2002.

J. K. Conner. Ecological genetics of floral evolution. In L. D. Harder and S. C. H. Barrett, editors, Ecology and Evolution of Flowers, pages 260–277. *Ecology and evolution of flowers*, 2006.

J. K. Conner and A. Sterling. Testing hypotheses of functional relationships: a comparative survey of correlation patterns among floral traits in five insect-pollinated plants. *American Journal of Botany*, 1995.

J. K. Conner and S. Via. Patterns of phenotypic and genetic correlations among morphological and life-history traits in wild radish, *Raphanus raphanistrum. Evolution*, 1993.

J. K. Conner, K. Karoly, C. Stewart, V. A. Koelling, H. F. Sahli, and F. H. Shaw. Rapid Independent Trait Evolution despite a Strong Pleiotropic Genetic Correlation. *The American Naturalist*, 178(4):429–441, Oct. 2011.

J. K. Conner, C. J. Mills, V. A. Koelling, and K. Karoly. Artificial selection on anther exsertion in wild radish, *Raphanus raphanistrum. Scientific Data*, 1:140027, Sept. 2014.

B. Devlin and N. C. Ellstrand. Male and female fertility variation in wild radish, a hermaphrodite. *The American Naturalist*, 136(1):87–107, 1990.

M. A. Dillies, A. Rau, J. Aubert, C. Hennequet-Antier, M. Jeanmougin, N. Servant, C. Keime, G. Marot, D. Castel, J. Estelle, G. Guernec, B. Jagla, L. Jouneau, D. Laloe, C. Le Gall, B. Schaeffer, S. Le Crom, M. Guedj, F. Jaffrezic, and on behalf of The French StatOmique Consortium. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA se- quencing data analysis. *Briefings in Bioinformatics*, 14(6):671–683, Nov. 2013.

P. K. Endress. The Immense Diversity of Floral Monosymmetry and Asymmetry Across Angiosperms. *The Botanical Review*, 78(4):345–397, 2012.

X. Feng, Y. Wilson, J. Bowers, R. Kennaway, A. Bangham, A. Hannah, E. Coen, and A. Hudson. Evolution of Allometry in *Antirrhinum*. *The Plant Cell*, 21(10):2999–3007, Dec. 2009.

J. Fornoni, M. Ordano, R. Pérez-Ishiwara, K. Boege, and C. A. Domínguez. A comparison of floral integration between selfing and outcrossing species: a meta-analysis. *Annals of Botany*, 11:mcv166–8, Nov. 2015.

A. Franzke, M. A. Lysak, I. A. Al-Shehbaz, M. A. Koch, and K. Mummenhoff. Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends in plant science*, 16(2):9–9, Feb. 2011.

B. J. Glover, C. A. Airoldi, S. F. Brockington, M. Fernández-Mazuecos, C. Martínez-Pérez, G. Mellers, E. Moyroud, and L. Taylor. How Have Advances in Comparative Floral Development Influenced Our Understanding of Floral Evolution? *International Journal of Plant Sciences*, 176(4):307–323, May 2015.

R. I. Greyson and V. K. Sawhney. Initiation and Early Growth of Flower Organs of Nigella and Lycopersicon: Insights from Allometry. *Botanical Gazette*, 133(2):184–190, Sept. 2015.

M. A. Grillo, C. Li, M. Hammond, L. Wang, and D. W. Schemske. Genetic architecture of flowering time differentiation between locally adapted populations of Arabidopsis thaliana. *New Phytologist,* pages n/a–n/a, Jan. 2013.

L. D. Harder and S. C. H. Barrett. Pollen Removal From Tristylous Pontederia Cordata: Effects of Anther Position and Pollinator Specialization. *Ecology*, 74(4):1059–1072, June 1993.

J. Hepworth and M. Lenhard. Regulation of plant lateral-organ growth by modulating cell number and size. *Current Opinion in Plant Biology*, 17:36–42, Feb. 2014.

Z.-L. Hu, J. Bao, and J. M. Reecy. CateGOrizer: A Web-Based Program to Batch Analyze Gene Ontology Classification Categories. *Online Journal of Bioinformatics*, 9(2):108–112, Sept. 2008.

L. Humeau and J. D. Thompson. The allometry of flower size dimorphism in dioecious Dombeya species on La Réunion. *Ecology Letters,* 4(3):221–228, May 2001.

R. E. Irwin and S. Y. Strauss. Flower Color Microevolution in Wild Radish: Evolutionary Response to Pollinator-Mediated Selection. *The American Naturalist*, 165(2):225–237, Feb. 2005.

R. E. Irwin, S. Y. Strauss, S. Storz, A. Emerson, and G. Guibert. The role of herbivores in the maintenance of a flower color polymorphism in wild radish. *Ecology*, 84(7):1733–1743, 2003.

Y.-M. Jeong, N. Kim, B. O. Ahn, M. Oh, W.-H. Chung, H. Chung, S. Jeong, K.-B. Lim, Y.- J. Hwang, G.-B. Kim, S. Baek, S.-B. Choi, D.-J. Hyung, S.-W. Lee, S.-H. Sohn, S.-J. Kwon, M. Jin, Y.-J. Seol, W. B. Chae, K. J. Choi, B.-S. Park, H.-J. Yu, and J.-H. Mun. Elucidating the triplicated ancestral genome structure of radish based on chromosome-level comparison with the Brassica genomes. *Theoretical and Applied Genetics*, pages 1–16, Mar. 2016.

H. Kitashiba and J. B. Nasrallah. Self-incompatibility in Brassicaceae crops: lessons for interspecific incompatibility. *Breeding Science*, 64(1):23–37, 2014.

H. Kitashiba, F. Li, H. Hirakawa, T. Kawanabe, Z. Zou, Y. Hasegawa, K. Tonosaki, S. Shirasawa, A. Fukushima, S. Yokoi, Y. Takahata, T. Kakizaki, M. Ishida, S. Okamoto, K. Sakamoto, K. Shirasawa, S. Tabata, and T. Nishio. Draft Sequences of the Radish (*Raphanus sativus L.*) Genome. *DNA Research*, 21(5):481–490, Oct. 2014.

T. Klinger, D. Elam, and N. C. Ellstrand. Radish as a Model System for the Study of Engineered Gene Escape Rates Via Crop-Weed Mating. *Conservation Biology*, 5(4):531–535, 1991.

E. M. Kramer. Understanding the Genetic Basis of Floral Diversity. *BioScience,* 57(6):479–487, June 2007.

B. A. Krizek. Making bigger plants: key regulators of final organ size. *Current Opinion in Plant Biology*, 12(1):17–22, Feb. 2009.

G. Kudo. Anther arrangement influences pollen deposition and removal in hermaphrodite flowers. *Functional Ecology*, 17(3):349–355, June 2003.

M. Love, S. Anders, and W. Huber. Package 'DESeq2' Manual, 2018.

S. J. Mazer and C. T. Schick. Constancy of population parameters for life-history and floral traits in *Raphanus sativus* L. II. Effects of planting density on phenotype and heritability estimates. *Evolution*, 45(8):1888–1907, 1991.

T. P. Miettinen and M. Björklund. Mitochondrial Function and Cell Size: An Allometric Relationship. *Trends in Cell Biology*, 27(6):393–402, June 2017.

Y. Mitsui, M. Shimomura, K. Komatsu, N. Namiki, M. Shibata-Hatta, M. Imai, Y. Katayose, Y. Mukai, H. Kanamori, K. Kurita, T. Kagami, A. Wakatsuki, H. Ohyanagi, H. Ikawa, N. Minaka, K. Nakagawa, Y. Shiwa, and T. Sasaki. The radish genome and comprehensive gene

expression profile of tuberous root formation and development. *Nature Publishing Group*, pages 1–14, June 2015.

G. D. Moghe, D. E. Hufnagel, H. Tang, Y. Xiao, I. Dworkin, C. D. Town, J. K. Conner, and S.-H. Shiu. Consequences of Whole-Genome Triplication as Revealed by Comparative Genomic Analyses of the Wild Radish *Raphanus raphanistrum* and Three Other Brassicaceae Species. *The Plant Cell*, 26(5):1925–1937, May 2014.

M. T. Morgan and J. K. Conner. Using genetic markers to directly estimate male selection gradients. *Evolution*, 55(2):272–281, 2001.

K. J. Niklas. Plant Allometry. The Scaling of Form and Process. *University of Chicago Press*, 1994.

J. Nishihiro and I. Washitani. Post-pollination process in a partially self-compatible distylous plant, Primula sieboldii (Primulaceae). *Plant Species Biology,* 26(3):213–220, 2011.

M. A. Nowak, M. C. Boerlijst, J. Cooke, and J. M. Smith. Evolution of genetic redundancy. *Nature*, 388(6638):167–171, July 1997.

J. G. Okie, V. H. Smith, and M. Martin-Cereceda. Major evolutionary transitions of life, metabolic scaling and the number and size of mitochondria and chloroplasts. *Proceedings of the Royal Society B: Biological Sciences,* 283(1831):20160611–8, May 2016.

A. E. Powell and M. Lenhard. Control of Organ Size in Plants Review. *Current Biology*, 22(9): R360–R367, May 2012.

J. C. Preston, L. C. Hileman, and P. Cubas. Reduce, reuse, and recycle: developmental evolution of trait diversification. *American Journal of Botany*, 98(3):397–403, Mar. 2011.

C. E. Ridley and N. C. Ellstrand. Evolution of enhanced reproduction in the hybrid-derived invasive, California wild radish (*Raphanus sativus*). *Biological Invasions*, 11(10):2251–2264, Dec. 2009.

H. F. Sahli, J. K. Conner, F. H. Shaw, S. Howe, and A. Lale. Adaptive Differentiation of Quantitative Traits in the Globally Distributed Weed, Wild Radish (*Raphanus raphanistrum*). *Genetics*, 180(2):945–955, Sept. 2008.

V. M. Savage, A. P. Allen, J. H. Brown, J. F. Gillooly, A. B. Herman, W. H. Woodruff, and G. B. West. Scaling of number, size, and metabolic rate of cells with body size in mammals. *Proceedings of the National Academy of Sciences of the United States of America,* 104(11):4718–4723, Mar. 2007.

D. W. Schemske. Adaptation and The Origin of Species. *The American Naturalist,* 176(S1):S4–S25, Dec. 2010.

F. Seyednasrollah, A. Laiho, and L. L. Elo. Comparison of software packages for detecting differ- ential expression in RNA-seq studies. *Briefings in Bioinformatics*, Dec. 2013.

D. R. Smyth, J. L. Bowman, and E. M. Meyerowitz. Early flower development in Arabidopsis. *The Plant Cell,* 2(8):755–767, Aug. 1990.

R. Snell and L. W. Aarssen. Life history traits in selfing versus outcrossing annuals: exploring the 'time-limitation' hypothesis for the fitness benefit of self-pollination. *BMC ecology*, 5:2, Feb. 2005.

P. S. Soltis, S. F. Brockington, M. J. Yoo, A. Piedrahita, M. Latvis, M. J. Moore, A. S. Chanderbali, and D. E. Soltis. Floral variation and floral genetics in basal angiosperms. *American Journal of Botany*, 96(1):110–128, Jan. 2009.

M. L. Stanton, A. A. Snow, and S. N. Handel. Floral evolution: attractiveness to pollinators increases male fitness. *Science*, 232(4758):1625–1627, June 1986.

T. Tian, Y. Liu, H. Yan, Q. You, X. Yi, Z. Du, W. Xu, and Z. Su. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic acids research*, 45(W1):W122–W129, May 2017.

A. Ushimaru and K. Nakata. Evolution of Flower Allometry and Its Significance for Pollination Success in the Deceptive Orchid Pogonia japonica. *International Journal of Plant Sciences,* 162 (6):1307–1311, July 2015.

C. J. Webb and D. G. Lloyd. The avoidance of interference between the presentation of pollen and stigmas in angiosperms II. Herkogamy. *New Zealand Journal of Botany*, 24(1):163–178, 1986.

T. D. Wu and S. Nacu. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, 26(7):873–881, 2010.

T. D. Wu and C. K. Watanabe. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, 21(9):1859–1875, May 2005.

M. F. Yanofsky. Floral meristems to floral organs: genes controlling early events in Arabidopsis flower development. *annualreviews.org*.