bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 # Patterns of population genomic diversity in the invasive Japanese
2 knotweed species complex
3

4 Acer VanWallendael[1*,†], Mariano Alvarez[2], Steven J. Franks[1]

5 [1]Fordham University, 441 E. Fordham Road, Bronx, NY 10458; [2]Duke University, 130 Science
6 Drive, Durham, NC 27708 *Corresponding Author: vanwaax07@gmail.com [†]Current address:
7 Michigan State University, 612 Wilson Road, East Lansing, MI 48824
8

9 **Running head**: Diversity in invasive clonal knotweed

10 **Keywords**: invasive species; genetic diversity; Japanese knotweed; clonal; polyploid

11

12

13

14

15

16

17

18

19

20

21

22

23

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24 **Abstract**

25 Premise: Invasive species are expected to experience a reduction in genetic diversity due
26 to founder effects, which should limit their ability to adapt to new habitats. Still, many invasive
27 species achieve widespread distributions and dense populations. This paradox of invasions could
28 potentially be overcome through multiple introductions or hybridization, both of which increase
29 genetic diversity. We conducted a population genomics study of Japanese knotweed (*Reynoutria*
30 *japonica*), which is a polyploid, clonally reproducing invasive species that has been notoriously
31 successful worldwide despite supposedly low genetic diversity.

32 Methods: We used Genotyping-by-Sequencing to collect 12,912 SNP markers from 88
33 samples collected at 38 locations across North America for the species complex. We used non-
34 alignment based k-mer hashing analysis in addition to traditional population genetic analyses to
35 account for the challenges of genotyping polyploids.

36 Results: Genotypes conformed to three genetic clusters, likely representing Japanese
37 knotweed, Giant knotweed, and hybrid Bohemian knotweed. We found that, contrary to previous
38 findings, the Japanese knotweed cluster had substantial genetic diversity, though it had no
39 apparent genetic structure across the landscape. In contrast, Giant knotweed and hybrids showed
40 distinct population groups. We did not find evidence of Isolation-by-Distance in the species
41 complex, likely reflecting the stochastic introduction history of this species complex. Among
42 species, we found no correlations between SNPs and several temperature- and precipitation-
43 based climatic variables.

44 Conclusions: The results indicate that clonal invasive species can show substantial
45 genetic diversity and can be successful at colonizing a variety of habitats without showing
46 evidence of local adaptation or genetic structure.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

47 **Introduction**

48 Understanding genetic changes accompanying population expansion is critical for predicting
49 ecological and evolutionary dynamics in colonizing species. Organisms colonizing a new area
50 often experience a type of population bottleneck known as a founder effect, in which a subset of
51 individuals starts a new population with reduced genetic diversity compared to the original
52 population (Mayr 1942). As the population expands in the new regions, serial founding events
53 may repeat across the landscape (Klopfstein et al. 2005; Excoffier & Ray 2008), magnifying the
54 effect. Resultant genetic drift during colonization may lead to reduced genetic diversity (Wright
55 1931; Dlugosch and Parker 2008; Excoffier and Ray 2008), which may in turn lead to loss of
56 beneficial traits as well as a reduction in the efficacy of selection (Frankham 1995; Peischl et al.
57 2013; Blackburn et al. 2015). Invasive species, typically defined as non-native organisms that
58 spread rapidly and displace natives, represent a magnified example of population expansion. In
59 what has been termed the "genetic paradox of invasion" (Frankham 1995; Allendorf & Lundquist
60 2003; Estoup et al. 2016), many introduced species, despite the predicted negative effects of the
61 loss of genetic diversity, are highly successful and become invasive. Invasive plants provide a
62 substantial threat to biological diversity, and their ecology and evolution has been studied for
63 decades (Butchart et al. 2010; Baker & Stebbins 1965), but researchers still know relatively little
64 about the genomics of invasive plants, including the extent and effect of population genetic
65 bottlenecks and the influence of factors such as clonal reproduction and hybridization on shaping
66 responses to selection, and invasion success, across the landscape (Lee 2002, Chown et al. 2015,
67 Allendorf and Lundquist 2003).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

68 The perennial, dioecious plant *Reynoutria japonica* has been described as one of the
69 "world's worst invasive alien species" for its aggressive nature and challenges in its removal
70 (Lowe et al. 2000). Since it is widespread in its introduced range, and reproduces through both
71 sexual and asexual means, it represents a useful system to test hypotheses about the impacts of
72 colonization on genetic diversity. The species is native to Japan and introduced to China,
73 Australia, Europe, and North America. In North America, the first known introduction occurred
74 in 1868 from a European source (Del Tredici 2017). Many subsequent introductions and
75 explosive growth have spread the species in high densities around temperate North America
76 (Barney 2006). Octoploid *R. japonica* forms a species complex with tetraploid *R. sachalinensis*
77 (F. Schmidt ex Maxim.) Nakai (Giant knotweed) to form hybrid hexaploid *R.* x *bohemica* Chrtek
78 & Chrtková (Bohemian knotweed; Bailey 2013). Both *R. sachalinensis* and hybrid *R.* x *bohemica*
79 are sympatric with *R. japonica* in North America, although their density is difficult to estimate
80 because many reports misidentify hybrids as *R. japonica* (Zika & Jacobson 2003; Gammon et al.
81 2007; Gaskin et al. 2014).

82 Notably, *R. japonica* has often been used as the quintessential example (Durka et al.
83 2005; Geng et al. 2007; Barrett et al. 2008; Herman & Sultan 2011) of an invasive species that is
84 successful over a wide range of habitats and environments despite its supposed partial or
85 complete lack of genetic diversity. *R. japonica* overwinters as a rhizome, and pieces of rhizome
86 that fragment will grow into clones of the parent plant (Hollingsworth & Bailey 2000; Barney
87 2006; Richards et al. 2008). In the native range, there is substantial chloroplast haplotype
88 diversity (Inamura 2000), but in the introduced range, *R. japonica* often has far less. For example,
89 a study using RAPDs (Randomly Amplified Polymorphic DNA) indicated that all *R. japonica* in
90 Great Britain is one female clone thought to be "one of the world's largest vascular plants"

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

91 (Hollingsworth & Bailey 2000). In North America, reports of genetic diversity have been mixed.
92 There is strong evidence through chloroplast markers that there are several haplotypes, and that
93 sexual reproduction is widespread (Grimsby et al. 2007; Grimsby & Kesseli 2010; Gammon et al.
94 2010, Gammon & Kesseli 2010). However, more recent AFLP (Amplified Fragment Length
95 Polymorphism) studies have found genetic diversity only in *R. sachalinensis* and *R.* x *bohemica*,
96 but none in *R. japonica* collected from both the Mid-Atlantic and Pacific Northwest regions
97 (Richards et al. 2012; Gaskin et al. 2014). This difference in conclusions may be either due to a
98 difference in sampling range or differences in the density and information content of the markers
99 used. Therefore, we predicted that an increased number of markers and broader sequencing over
100 a wide geographic range would help to clarify the distribution of genetic diversity present in
101 North America.
102 The adaptive consequences of asexual reproduction are crucial to understanding the
103 Japanese knotweed invasion. As colonizers, selfing and clonal plants possess several inherent
104 advantages over obligate outcrossers, which could help explain why the capacity for uniparental
105 reproduction is more common in invasive species than in those that do not become invasive
106 (Pannell 2015; Razanajatovo & van Kleunen 2016). These advantages include a lack of need for
107 other mates (Baker 1955), genetic continuity across generations that helps provide continued
108 success in a given environment (Baker & Stebbins 1965), and reduced gene flow, which fosters
109 local adaptation (Kawecki & Ebert 2004; Baker 1955; Baker & Stebbins 1965). But a
110 disadvantage of asexual reproduction is that genetically limited populations that may be less able
111 to adapt to novel biotic and abiotic stress (Crow & Kimura 1965; Verhoeven et al. 2010). It is
112 well-established that *R. japonica* typically reproduces asexually, while *R. sachalinensis* and
113 hybrids tend to spread through outcrossed seed (Bailey 2013). However, the population-level

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

114 impacts of these reproductive systems on the distribution of genetic diversity have not yet been
115 assessed.
116 This study examines the population and landscape genetics of the Japanese knotweed
117 species complex in North America with the goal of understanding the factors that have permitted
118 its high fitness despite founder effects. We used Genotyping-by-Sequencing (GBS) for 88
119 samples at 38 collection sites to capture 12,912 SNPs. Specifically, we sought to determine
120 whether or not genetic diversity has been limited in Japanese knotweed due to founder effects
121 combined with clonal propagation. Further, we hypothesized that genetic diversity would be
122 partitioned differentially within the species complex, with higher genetic diversity in *R.*
123 *sachalinensis* and hybrids. Finally, we examined the geographic dispersal of genotypes and their
124 correlation with climatic variables to test whether founder effects have limited the adaptive
125 potential of knotweed in the invaded range. The results of this work will provide insight into
126 population genetic processes of species that can successfully colonize a wide variety of habitats.

127 ## Materials and Methods

128 ### Study design

129 We structured our sampling to examine diversity at multiple geographic scales to
130 understand the landscape and microgeographic structure of populations. We did not make *a-*
131 *priori* distinctions between hybrid species in the field to avoid biased sampling. We collected
132 fresh leaf or rhizome tissue from 38 sites in the eastern portion of the United States (Figure 1;
133 Table S2). When possible, fresh leaves were collected and frozen within one hour at -20°C until
134 extraction. At more distant sites, we collected rhizomes into a cooler, then germinated them in the
135 greenhouse to obtain fresh leaves. Within sites, we collected several individuals and sequenced at

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

136 least two individuals per site whenever possible. As *R. japonica* spreads mainly via rhizomes,
137 groups of "individuals" may in fact be one organism. Since distinguishing individuals is
138 impossible barring excavation of the rhizome network, we followed the suggestions of Richards
139 et al. (2008) and collected only from rhizomes spaced 10 m apart. For Midwest and West Coast
140 samples, leaves were collected and dried by citizen scientists then sent by mail. During
141 collection, we assigned each stand to one of four habitat types: riparian, roadside, forest edge, or
142 lacustrine. Riparian, lacustrine, and roadside habitats were all within 50 m of a river, lake, or
143 road, respectively. All other sites were found at the border (forest edge) between a wooded area
144 and an anthropogenic landscape. We recorded the sex of each individual based on floral
145 characteristics and used the protocol of Zika & Jacobson (2003) to assign plants to a species.

146 **Extraction and sequencing**

147 We extracted DNA from leaf tissue using a FastPrep homogenizer and a Qiagen Plant
148 Mini kit (Qiagen, Hilden, Germany). We quantified DNA concentration using a Qubit
149 fluorometer and checked quality by running the samples on a 1% agarose gel (Invitrogen;
150 Carlsbad, CA). We sent samples to the Cornell Institute of Biotechnology for library preparation
151 and sequencing. Genotyping-By-Sequencing (GBS) was performed according to the protocol
152 established by Elshire (2011). Briefly, this process fragments DNA using a restriction enzyme
153 (EcoT22I), then ligates barcoded primers to the restriction ends. The barcodes allow samples to
154 be distinguished after the next step, multiplexed PCR. The facility used an Illumina HiSeq to
155 sequence 100 bp single-end reads for all samples.

156 **Genetic diversity**

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

157 We used the STACKS (version 2.40) platform to process sequences and call SNPs
158 (Catchen et al. 2011). We first de-multiplexed the reads and removed barcodes using the
159 *process_radtags* script. Since knotweed lacks a reference genome, we used the non-referenced
160 aligned SNP-calling pipeline, *denovo_map*, in STACKS. SNP calling in polyploids is
161 complicated by the presence of several possible alleles at polymorphic loci (reviewed by Blischak
162 et al. 2017, Box 1). To maximize the likelihood of calling accurate SNPs, we used a method
163 similar to that used to call SNPs in tetraploid sturgeon (Anderson et al. 2017). We collected only
164 loci that had a minimum depth of eight (-m 8), a maximum of two mismatches between aligned
165 sequences within an individual (-M 2), and a maximum of one mismatch when comparing to the
166 catalog (-n 1). Since STACKS does not explicitly allow for polyploidy, this method concatenates
167 homeologs. This results in loss of loci due to allele dosage uncertainty, which is the uncertainty
168 over whether copy number is due to sequencing depth or multiple homologous chromosomes
169 (Blischak et al. 2017). To optimize filtering parameters, we ran several iterations of the
170 *populations* module to sequentially test parameters to give the greatest number of polymorphic
171 loci present in 80% of different populations (Paris et al. 2017). We ultimately filtered loci for
172 only SNPs that were present in at least 60% of samples with a coverage above 10x, using a
173 minimum minor allele frequency of 0.05, and a maximum observed heterozygosity of 0.9. This
174 resulted in 12,912 SNPs (sequences publicly available on the NCBI SRA: BioProject
175 PRJNA574173). Genetic diversity estimates of expected heterozygosity (H_e), nucleotide diversity
176 (π), and percent polymorphic loci were exported from STACKS.
177 Since genotyping error can cause overestimation of genetic diversity, we checked
178 diversity estimates against a subset of high-confidence SNPs. We re-ran the STACKS SNP
179 calling pipeline with very strict parameters to get a shorter list of 265 SNPs. This subset had

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

180 12.7% missing data with average coverage over 60x in all individuals (minimum coverage of 20x
181 at all loci). The high coverage means that it is extremely unlikely that any of these loci represent
182 genotyping errors. We then compared genetic diversity between the entire SNP set to high-
183 confidence loci to account for genotyping error.
184
185 **Population structure**
186 We calculated genetic distance between individuals from STACKS-generated SNPs using
187 the *dist.genpop* function in *adegenet*, and made the matrix Euclidean using *cailliez* in *ade4*
188 (Jombart 2008; Dray & Dufour 2007). We performed Principal Coordinate Analysis (PCoA) on
189 both distance matrices using the *dudi.pco* function in *ade4* (Dray and Dufour 2007). PCoA is
190 often more reliable than Principal Component Analysis (PCA) for GBS data because the latter is
191 more influenced by missing data (Legendre and Legendre 2012). However, to corroborate these
192 patterns, we also performed Principal Component Analysis (PCA) using *pcadapt* (Luu and Blum
193 2017).
194 Since SNPs called from polyploid samples may be unreliable, we compared our results to
195 an alternate method for measuring genetic distance known as the MinHash technique. The
196 MinHash technique is a form of probabilistic locality-sensitive hashing (Indyk & Motwani 1998)
197 that reduces the complexity of a dataset to a representative 'sketch' (Ondov et al. 2016). In the
198 sketching process, *Mash* converts randomly sampled *k*-mers (sequences of length *k*) from raw
199 sequence reads into computationally efficient 'hashes', many of which form a sketch. Sketches
200 can be thought of as genetic summaries and can be compared to find a reliable estimate for the
201 mutation rate between two sequences with relatively low computational load (Ondov et al. 2016).
202 This process takes advantage of the fact that hashes can be more rapidly compared than

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

203 sequences aligned and, due to random sampling, the degree to which *Mash* can accurately sketch
204 a given sequence scales with the size of the sketch, rather than the size of the genome (Ondov et
205 al. 2016). Though some precision is lost, genetic distances can be estimated without many of the
206 assumptions of other programs that handle genetic data. In particular, *Mash* does not assume
207 diploidy within samples. We sketched each individual using a *k*-mer size of 27 (-k 27) and
208 1,000,000 hashes (-s 1000000) per individual, omitting *k*-mers that appeared fewer than five
209 times (-c 5) before calculating *Mash* distance. *Mash* produced a distance matrix that we analyzed
210 in parallel to the outputs from STACKS. To test the relationship between the *Mash* distance
211 matrix and our Nei's distance matrix, we ran a Mantel test using 1,000 permutations.

212 To visualize phylogenetic relationships, we produced a neighbor-joining (NJ) tree using
213 *phangorn* 2.3.1 (Shliep 2010). We optimized this tree using both maximum parsimony and
214 maximum likelihood (ML) methods through the *parsimony* and *optim.pml* functions in *phangorn*.
215 We generated several trees that were compared through log likelihood. We compared models
216 using both AICc and BIC. Bootstrap values were produced by 100 iterations.

217 We used the program ADMIXTURE 2.40 to detect structure and admixture between
218 populations (Alexander et al. 2009). ADMIXURE, similarly to the commonly used
219 STRUCTURE software (Pritchard et al. 2000), uses multi-locus genotype data to detect
220 populations by assigning individuals to inferred ancestral populations. We ran ADMIXTURE for
221 all values of K between one and thirty-eight (the number of sites sampled) using the default
222 parameters.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

223 **Landscape variables**

224 We tested for evidence of Isolation-by-Distance (IBD) using distance-based redundancy
225 analysis (dbRDA) implemented with the *capscale* function in the *vegan* package (Oksanen et al.
226 2017) in R. In this analysis, we decomposed the geographic distance matrix into principal
227 components, then tested how the variance in the genetic distance matrix is explained by the
228 spatial eigenvectors using ANOVA (Legendre & Legendre 2012). The dbRDA analysis also
229 allowed testing of different factors to explain variation in the genetic distance matrix, including
230 site, species, sex, and habitat. To determine significance, we used PERMANOVA (Permutational
231 multiple analysis of variance). PERMANOVA is used to compare multivariate groups and tests
232 the null hypothesis that the centroids of the groups are equivalent. Unlike traditional MANOVA,
233 PERMANOVA measures significance by comparing the *F* test result to random permutations of
234 samples from each group (Oksanen et al. 2017). To choose the best fitting model, we used
235 stepwise model selection in the *ordistep* function of *vegan* (Oksanen et al. 2017). The best model
236 was based on a maximum of 50 steps of addition and removal of variables to determine best fit,
237 using permutation *P*-values as an alternative to AIC (Oksanen et al. 2017). In conjunction, we
238 measured the explanatory power of all nineteen Worldclim environmental variables (Table S1)
239 on our two distance matrices using Latent Factor Mixed Models (LFMM) implemented in the
240 *LEA* package in R (Frichot & Francois 2015). LFMMs provide a compromise between power and
241 error rate that is relatively conservative when testing environmental associations (Frichot et al.
242 2013; Villemereuil et al 2014). Their power lies in the simultaneous testing of environmental
243 correlations while estimating the hidden effect of population structure, so that these are not
244 confounded. The SNPs associated with the environmental variables were determined based on
245 their z-score. Z-score is calculated by the Gibbs sampler algorithm run for 50,000 sweeps after a

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

246 burn-in period of 10,000 sweeps. The threshold for the z-scores was determined after a
247 Bonferroni correction of α = 0.01. Loci with z-scores > 4.0 and $p < 10^{-5}$ were considered
248 significantly associated.

249 **Results**

250 **Genetic diversity**

251 *Diversity measures*

252 We found substantial genetic diversity in the populations of *R. japonica* we investigated
253 (π = 0.0024, SE = 0.0001; Table 1). Genetic diversity was greater in *R. sachalinensis* and *R.* x
254 *bohemica* than in *R. japonica* according to heterozygosity measurements (Table 1). Percent
255 polymorphic loci was greatest in the hybrid, and lower for both parent species. Conversely, the
256 largest number of private SNPs were found in the *R. sachalinensis* population. The lowest
257 number of private alleles were found in *R.* x *bohemica*.

258 We also measured genetic diversity in a subset of high-confidence SNPs. We found that
259 all measures of diversity were comparable in the subset (Table S3). We observed nucleotide
260 diversity of 0.0021 ± 0.0002 for *R. japonica,* 0.0023 ± 0.0003 for *R.* x *bohemica*, and 0.0030 ±
261 0.002 for *R. sachalinensis*. While these estimates were lower than what we found for the full SNP
262 set, we found the same pattern, with *R. japonica* the least diverse, and *R. sachalinensis* the
263 highest.

264 **Population structure**

265 Using several different approaches, we found little evidence of genetic structure among
266 populations of each of species of knotweed we investigated, although there were genetic

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

267 differences among the species. We first compared genetic distance matrices produced by two
268 independent methods, SNP-based genotyping and *Mash* distance. We observed a close
269 correlation between the SNP-based Nei's distance matrix and the *Mash* distance matrix using a
270 Mantel test (1000 permutations, $r = 0.625$, $p = 0.001$). Next, we performed principal coordinates
271 analysis to summarize variation in both matrices. The first and second principal coordinate axes
272 of the PCoA explained 22% and 5.8% of the variation for Nei's distance and 19% and 3.7%,
273 respectively, for *Mash* distance (Figure 3a,b). The first axis separated samples based largely on
274 species within the hybrid complex. The second and third axes exposed differences in genetic
275 diversity within species for both matrices. However, in the *Mash* distance matrix, the second and
276 third PCs were switched. We saw strong structure based on species and similar clustering (Figure
277 3a,b) for both matrices. Most sites contained only *R. japonica* and therefore, clustered together,
278 revealing very little population-based structure (Figure 3c,d). The sites containing *R.*
279 *sachalinensis* were distinct from *R. japonica* samples and from each other. The sites containing
280 *R.* x *bohemica* were the most variable on PC1 and were intermediate between the parental species
281 (Figure 3a,b). All of the observed patterns were consistent between PCoA representations of both
282 distance matrices (Figure 3a,b). The PCoA representation of *R. japonica* analyzed alone showed
283 a cluster of samples on the positive side of PC1 that may represent clonal samples (Figure 3c).
284 However, most sites separated on both PCs and sites did not form distinct groups.
285 Our ML trees mostly corroborated the results we found in the principal coordinate
286 analysis. In particular, *R. japonica* populations were not differentiated, but both *R. sachalinensis*
287 and *R.* x *bohemica* populations showed significant site-based differentiation (Figure 4a).
288 Bootstrap values were low for most branches within the *R. japonica* clade, but high within the
289 other clades. Notably, the tree highlights a strong isolation of two particular samples at the top of

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

290 the phylogeny, which were male *R.* x *bohemica* within a site containing mostly female *R.*
291 *japonica*.
292 We examined ADMIXTURE values of K from 1-38 and found the best support for K = 2
293 (Figure 4b, c). Consistent with other analyses, we saw distinction of *R. japonica* and *R.*
294 *sachalinensis* and admixture in sites that contained *R.* x *bohemica* hybrids. This is most apparent
295 when K = 2. The Cornell population is grouped with *R. sachalinensis* at K = 2, but shown to be
296 more similar to the *R.* x *bohemica* group when K is increased to 3. The homogenous block of *R.*
297 *japonica* was not separated for any level of K less than 10. At levels above 10, individuals were
298 assigned to differing inferred ancestral populations, but these did not correspond to collection
299 site. Analysis of *R. japonica* in isolation did not return different patterns than when analyzed in
300 conjunction with other species.

301 **Landscape variables**

302 We did not find evidence for Isolation-by-Distance (IBD) because there was no
303 relationship between geographic distance and genetic distance, as shown by dbRDA ($F = 0.907$,
304 $p = 0.679$; Table 2, Figure 4d). This pattern was not affected by the distance measure used, as
305 there was similarly low explanation of the variation in the *Mash* genetic distance by geographic
306 distance ($F = 1.03$, $p = 0.387$; Table 2). However, the combination of factors that best explained
307 genetic variation in PERMANOVA differed between measures of genetic distance. While both
308 models included species and site, Nei's distance was partially explained by habitat, and *Mash*
309 distance by sex, though the marginal effects of each of these factors were weak relative to species
310 and site (Table 2).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

311 We examined correlations with environmental variables using Latent Factor Mixed
312 Models. However, loci were not significantly correlated with any of nineteen climatic variables
313 (after correction for multiple testing; all values of $p > 0.9$).
314

315 Discussion

316 Overview

317 In this study, we examined one of the worlds' most successful and widespread weeds
318 using next-generation sequencing and genomic analyses. In contrast to the expectation that
319 invasive species in general should have low genetic diversity because of founder effects, and also
320 contrary to prior studies indicating the near absence of genetic diversity in this group, we found
321 substantial genetic diversity in the Japanese knotweed species complex in the North American
322 populations we sampled. However, there was little population structure, and the distribution of
323 genetic diversity was not correlated with climatic or geographic factors in *R. japonica*. Our study
324 highlights the importance of careful assessment of genetic diversity and consideration of the role
325 multiple introductions and hybridization can take in species invasions. This work also provides
326 an illustration of how contemporary analytical techniques can be employed to evaluate
327 population structure of polyploid species lacking a reference genome.

328 Differences in genetic diversity between hybrid species

329 Due perhaps to its ubiquity and invasiveness, *R. japonica* has been the focus of several
330 genetic studies over the last two decades in North America. While sexual reproduction was
331 known to occur in *R. japonica* populations, most studies assumed that the majority of
332 reproduction occurred through asexual means (Forman & Kesseli 2003; Richards et al. 2012).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

333 Rather than completely a simple binary clonal versus non-clonal designation, researchers have
334 recognized that the genetic definition of clonality is a continuum of similarity (Bailleul et al.
335 2016). Strategies for dividing samples into clonal groups, or multi-locus lineages, usually involve
336 designating a minimum threshold for genetic distance that corresponds to an estimate of
337 sequencing error and somatic mutation (Kamvar et al. 2015). In the absence of prior knowledge
338 of this threshold, this strategy relies on a frequency distribution of genetic distances that often
339 shows a peak near zero designating clonal lineages (Meirmans & van Tienderen 2004). We can
340 reject the hypothesis that North American *R. japonica* are entirely clonal, since even with a
341 conservative threshold for minimum clonal distance we would observe several clonal linages;
342 most samples showed moderate genetic distance (D; range 0.014-0.201; median: 0.094). Studies
343 using SNP markers have observed similar genetic distances of between 0.02 and 0.19 for
344 subpopulations of sexual *Solanum lycopersicum* (Sim et al. 2015). Further, the level of genetic
345 diversity we measured ($\pi = 0.0024$) is comparable to what has been seen in other polyploid
346 species (Cornille et al. 2016), and much higher than what has been measured in clones
347 (Gutekunst et al. 2018). Because both genetic distances and overall diversity were higher than
348 would be expected for purely clonal lineages, we suggest that stands of *R. japonica* often contain
349 related individuals in addition to clones.
350 Precise estimation of genetic diversity is problematic in polyploid organisms because
351 estimated heterozygosity can be inflated due to homologous chromosomes yielding homeologous
352 sequences (Blischak et al. 2017). In addition, taxa that undergo whole genome duplication often
353 then undergo diploidization, a process by which homeologs are lost and diploid inheritance is
354 resumed (Barker et al. 2016). If the SNPs are mostly clustered in non-coding regions or on
355 silenced homeologous chromosomes, different genotypes may be functionally identical

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

356 (Mandáková et al. 2017). Further, somatic mutations occur often in plants, and can play an
357 important role in evolution, especially in clones (Whitham & Slobodchikoff 1981; Schoen &
358 Schultz 2019). While it is conceivable that some of the genetic diversity we observe is due to
359 somatic mutations within clones, it is unlikely to be a major source of genetic diversity in
360 introduced areas, given the relatively short time since *Reynoutria spp.* were introduced to North
361 America. Even if somatic mutations did contribute to genetic diversity in this complex, this
362 would not contradict our finding of substantial genetic diversity, but would provide one
363 mechanism for the generation of this diversity.

364 We took several measures to ensure that genetic diversity estimates we report are not due
365 to artifacts of sequencing or data analysis, including conducting the analyses with a subset of
366 high-confidence SNPs and confirming the results with *Mash* analysis. The congruence of the
367 analysis on the high-confidence SNP data and the full dataset indicates that the patterns we found
368 are unlikely to be due to artifacts or error. While the values we report for π and H_s are not
369 necessarily comparable to those for a diploid species, they represent a useful estimation of
370 relative genetic diversity. *Mash* estimation is less influenced by ploidy than alignment-based
371 SNPs, so it is a useful benchmark to show that within-population diversity is substantial. Thus, it
372 is unlikely that polyploidy prevented this study from correctly identifying the pattern of genetic
373 diversity.

374 There is a somewhat surprising pattern in *Reynoutria* species of higher-ploidy species
375 exhibiting lower genetic diversity (Hollingsworth et al. 2000; Richards et al. 2012). Even if they
376 are not entirely clonal, octoploid *R. japonica* show less nucleotide diversity than either tetraploid
377 *R. sachalinensis* or hexaploid *R.* x *bohemica.* Polyploid individuals have inherently higher allelic
378 diversity than diploids, since a duplicated genome allows twice the potentially expressed alleles.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

379 However, at the population level polyploid populations often have lower genetic diversity owing
380 to life history constraints. In a striking recent example, a triploid genotype of marbled crayfish
381 (*Procambarus virginalus*) have been successful in invading ecosystems worldwide (Gutekunst et
382 al. 2018), despite complete genetic identity. The pattern is also evident in native species
383 colonizing new areas. Native Pacific Northwest hawthorns (*Crataegus spp.*) show high genetic
384 diversity throughout a species complex, but several clonal polyploid lineages have colonized
385 wider ranges than their diploid ancestors (Coughlan et al. 2017). Similarly, octoploid *R. japonica*
386 shows both higher asexual reproduction and wider range than tetraploid *R. sachalinensis*. There
387 is a well-established association between polyploidy and asexual reproduction in both plants and
388 animals (Otto & Whitton 2000; Neiman et al. 2014). However, the reasons for the repeated
389 pattern of range expansion following asexual polyploidization are not completely understood.
390 One possible explanation is that polyploidization allows a species access to a new niche (Baniaga
391 et al. 2020), and asexual reproduction allows the neopolyploid a means to reproduce in the
392 absence of mates. The polyploid history of *R. japonica* has not yet been studied, so a relationship
393 between niche divergence and polyploidization cannot be determined. Future studies may be able
394 to exploit this polyploid series to better understand the importance of ploidy to plant colonization.

395 **Population Structure**

396 Population structuring, like diversity, differed between *Reynoutria* species. The difference
397 between species was unsurprisingly the strongest signal of differentiation, but there was also
398 some structuring by collection site for *R. sachalinensis* and *R.* x *bohemica*. Within *R. japonica*,
399 we saw no evidence of population structuring (Figure 3). While this could be evidence of
400 panmixia, wherein gene flow is high between all sites, the wide geographic range of sites means
401 that this explanation is highly unlikely (Waples & Gaggiotti 2006). Instead, we hypothesize that

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

402 the *R. japonica* samples we analyzed were derived from a homogenous source population, and
403 have not yet genetically differentiated within the invaded range. This interpretation is
404 corroborated by undetectable Isolation-by-Distance (IBD) in these populations, as well as the
405 stochastic nature of *R. japonica* introductions to North America (Barney 2006). Within the hybrid
406 and *R. sachalinensis*, there was strong structuring, as was seen in other studies (Grimsby et al.
407 2009; Richards et al. 2012; Gaskin et al. 2014). Depending on the analysis used, we saw slightly
408 differing patterns. While PCoA clearly distinguishes Cornell and Maine from all other sites, the
409 ML tree included several *R.* x *bohemica* from CT as closely related to Cornell and Maine (two
410 samples at the top of the tree). One possible explanation is that these CT samples are introgressed
411 backcrosses between *R.* x *bohemica* and *R. sachalinensis*, as has been found previously in
412 American populations (Gammon et al. 2010).
413 Another notable pattern in population structure is the confounding factor of sex.
414 *Reynoutria* are subdioecious, with females and hermaphrodite individuals that rarely produce
415 seeds (*i.e.* function as males; Forman & Kesseli 2003). The clonal UK population of *R. japonica*
416 consists entirely of females, though male *R. sachalinensis* and *Fallopia balduschiana* often
417 provide pollen for hybridization (Hollingsworth & Bailey 2000; Bailey 2013). In contrast, male
418 *R. japonica* are present (if rare) in North America (Forman & Kesseli 2003). Male *R.*
419 *sachalinensis* and *R. bohemica* are far more common, and account for the entire Maine
420 population as well as two samples which show great divergence from the CT population (Figure
421 4a,b). Biased sex ratios impact N_e and therefore may have wide-ranging effects on estimation of
422 genetic diversity. Biased sex ratios should be selected against, since the rarer sex has a
423 comparative fitness advantage (Fisher 1958). Biased ratios can persist due to several factors,
424 including parthenogenesis as in several vertebrates (Lynch 1984), *Wolbachia* infection in insects

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

425 (Stouthamer et al. 1999), and partial reproductive isolation variation in plants (Barnard-Kubow &
426 Galloway 2017), to name a few. In the case of knotweed, it may simply be the case that the
427 rhizomes that were moved to Europe and North America were more commonly female than male,
428 and that populations have not have had time to equalize.

429 **Landscape variation**

430 Unlike several other population genomic studies in invasive species (Cornille et al. 2016;
431 Trumbo et al. 2016; Combs et al. 2018), we did not find evidence of Isolation-By-Distance
432 (IBD). While the introduction history of *Reynoutria spp.* is complicated, lack of IBD for a
433 widespread species with no clear dispersal boundaries is fairly surprising. PERMANOVA
434 showed collection site-based structure, which indicates that there is some geographic component
435 to genetic structure. However, the fact that the populations do not show IBD indicates that these
436 populations are scattered across the landscape, likely due to a stochastic introduction history
437 rather than linear spread. In addition, we also observed no correlations between genetic variation
438 and climatic variables. There are several possible explanations for this finding. The LFMM
439 method uses a conservative correction for multiple testing that may result in values below the
440 significance limit (Frichot et al. 2013). However, PERMANOVA indicates that *R. japonica* has
441 been introduced stochastically across the landscape, and this may result in little geographic
442 differentiation. While we found evidence of sexual reproduction, frequent clonal propagation of
443 *Reynoutria spp.* likely plays a strong role in both preventing IBD and minimizing environmental
444 correlations (Reusch et al. 2000). Research on clonal invasive cogongrass (*Imperata cylindrica*)
445 showed similar lack of IBD, presumably due to limitations in diversity (Burrell et al. 2015). In
446 the *R. japonica* native range, there is predictable species separation, but no information on

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

447 hybrids. *R. sachalinensis* grows mostly in the northern Sakhalin island in Russia and Hokkaido,
448 the northernmost island in Japan (Inamura 2000). Although the PERMANOVA did not detect an
449 overall effect of distance, we only found *R. sachalinensis* in the North (Figure 1), suggesting a
450 pattern that might be detected with wider sampling. The lack of landscape-scale patterns may be
451 a product of both life history and invasion history. In outcrossing populations, adaptation along a
452 linear gradient promotes differentiation (Rousset 1997, Vekemans & Hardy 2004), but
453 populations that show clonal reproduction diverge genetically much more slowly during spread.
454 As a compounding factor, multiple introductions of *Reynoutria spp.* (Barney 2006) have likely
455 led to a complicated mosaic of genetic diversity across the landscape that is challenging to
456 untangle at this level of resolution.
457

458 **Population genetics of polyploids**

459 There are several distinct challenges of inferring population genetic patterns from
460 polyploid data. In general, sequencing to sufficient depth to capture allele variation is expensive,
461 and large genome size for many polyploids, including *R. japonica*, further complicates matters. A
462 locus that is partially heterozygous creates uncertainty in dosage, which complicates accurate
463 inference of allele frequencies (Meirmans et al. 2018). In addition, most population genetic
464 theory is built around outcrossing diploid models that may or may not apply to clonal polyploids,
465 including simple measurements such as fixation index (F_{ST}; Dufresne et al. 2014; Meirmans et al.
466 2018). At each stage of analysis (genotyping, estimation of population structure, and landscape
467 associations), we used parallel analysis techniques to account for some of this uncertainty, and
468 relied on relative measures of genetic diversity whenever possible to avoid bias from genotyping
469 uncertainty. Across analyses, we found broadly similar results, supporting our final conclusions.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

470 Whole-genome sequencing is becoming economically feasible for a wider array of
471 polyploid species, so may soon replace reduced-representation methods (e.g. Edger et al 2019).
472 However, reduced-representation methods remain an excellent means of capturing variation in
473 many markers for large sample sizes. To understand coarse patterns in data, *Mash* and similar k-
474 mer hashing techniques may be particularly useful due to the few assumptions required and the
475 low computational load (Ondov et al. 2016; Viochek & Weigel 2019). However, it lacks many of
476 the data-cleaning steps of other genotyping platforms and therefore may be prone to
477 overestimating genetic diversity. The *STACKS* platform has numerous well-curated tools for
478 many applications. For projects with many markers, however, a high-performance computer is
479 needed. Population genetic studies in polyploid species will always face challenges, but
480 bioinformatic tools are increasingly accommodating higher ploidy levels.
481
482 **Conclusions**
483 Japanese knotweed populations in North America display a complex genetic structure that
484 differs from what has been seen at smaller scales with other genetic tools. Unlike what has been
485 traditionally understood for North America, and what has been seen in Great Britain, there is
486 substantial genetic diversity within populations of knotweed. While clonal spread is undoubtedly
487 important to the invasive success of knotweed, it is evident that sexual reproduction has also
488 occurred during its rapid spread due to the genetic variation that exists with sites, although
489 somatic mutation could have potentially also played some role in creating diversity. The lack of
490 evidence for IBD in knotweed populations in the present study is likely linked to its stochastic
491 introduction throughout the late 19th and early 20th centuries, and its frequent clonal propagation
492 and less frequent sexual reproduction. We provide evidence that knotweed has avoided the

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

493 genetic paradox of invasions (rapid spread despite loss of genetic variation) by two means. First,
494 populations that were assumed to be clonal harbor levels of heterozygosity that indicate there has
495 not been complete loss of diversity in introduced populations, either because of sexual
496 reproduction or retained polyploid diversity. While these populations have not differentiated
497 across the landscape, phenotypic variability may not be limited by lack of genetic variation
498 (Richards et al. 2012). Second, hybridization with *R. sachalinensis* occurs readily, so the species
499 complex as a whole can be said to avoid the genetic paradox through hybridization. Since
500 polyploidy can play a role in both harboring heterozygosity and allowing hybridization, we
501 expect that it has influenced knotweed invasion. Finally, we have shown that GBS data can be
502 viable for non-model polyploids by comparing our results between two fundamentally different
503 methods for measuring genetic differentiation. Taken together, our results provide insights into
504 the roles of factors such as clonality and hybridization in shaping the population genomics and
505 success of an invasive species, and more broadly help to inform our understanding of the process
506 of colonization.
507

508 **Acknowledgements:** We thank D. Fogarasi, S. Davey, G. Diaz, and J. Goehl for field assistance;
509 A. VanWallendael, P. VanWallendael, K. Ansaldi, J. Huffstetler, B. Wade, Dr. N. Bauer, Dr. J.
510 Weber, and Dr. N. Muth for sending samples; D. Fogarasi and J. Park for lab assistance; M.
511 Combs and E. Puckett for assistance with data analyses; and Dr. Jason Munshi-South, Dr. John
512 Wehr, Dr. James Lewis, Dr. Beth Ansaldi, Dr. Michael Sekor for helpful comments on the
513 manuscript. We acknowledge the following field stations and reserves: Louis Calder Center,
514 Goodwin College, USFS, Hubbard Brook Experimental Forest, Mohonk Reserve, Blandy
515 Experimental Farm, Wells Reserve, Cornell University, New York Botanical Garden, Claytor

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

516 Nature Study Center. We thank the Fordham Graduate School of Arts and Sciences, Fordham
517 Graduate Student Association, and Louis Calder Center for funding.

518

519 **Author Contributions**: AV designed the study, collected and analyzed data and wrote the
520 manuscript. MA analyzed data and wrote the manuscript. SJF contributed to study design and
521 writing of the paper.

522

523 **Data Availability:** Sequences have been deposited into the NCBI Short Read Archive (SRA) as
524 BioProject PRJNA574173

525

526 **Literature Cited:**

527 Abbott, R. J. (1992). Plant invasions, interspecific hybridization and the evolution of new plant
528 taxa. *Trends in Ecology & Evolution*, *7*(12), 401–405.

529 Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in
530 unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

531 Allendorf, F. W., & Lundquist, L. L. (2003). Introduction: population biology, evolution, and
532 control of invasive species. *Conservation Biology*, *17*(1), 24–30.

533 Anderson, E. C., Ng, T. C., Crandall, E. D., & Garza, J. C. (2017). Genetic and individual
534 assignment of tetraploid green sturgeon with SNP assay data. *Conservation Genetics*, 1–12.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

535 Ayres, D. R., Zaremba, K., Sloop, C. M., & Strong, D. R. (2008). Sexual reproduction of
536 cordgrass hybrids (*Spartina foliosa* x *alterniflora*) invading tidal marshes in San Francisco
537 Bay. *Diversity and Distributions*, *14*(2), 187–195.

538 Bailey, J. (2013). The Japanese knotweed invasion viewed as a vast unintentional hybridisation
539 experiment. *Heredity*, *110*(2), 105–110.

540 Baker, H. G. (1955). Self-compatibility and establishment after ‘long-distance’ dispersal.
541 *Evolution*, *9*(3), 347–349.

542 Baker, H. G., & Stebbins, G. L. (1965). genetics of colonizing species, proceedings. In
543 *International Union of Biological Sciences Symposia on General Biology 1964: Asilomar,*
544 *Calif.)*. Academic Press.

545 Bailleul, D., Stoeckel, S., & Arnaud□Haond, S. (2016). RClone: a package to identify
546 MultiLocus Clonal Lineages and handle clonal data sets in r. *Methods in Ecology and*
547 *Evolution*, *7*(8), 966-970.

548 Baniaga, A.E., Marx, H.E., Arrigo, N. and Barker, M.S., 2020. Polyploid plants have faster rates
549 of multivariate niche differentiation than their diploid relatives. *Ecology letters*, *23*(1),
550 pp.68-78.

551 Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z., & Levin, D. A. (2016). On the relative
552 abundance of autopolyploids and allopolyploids. *New Phytologist*, *210*(2), 391–398.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

553 Barnard□Kubow, K. B., & Galloway, L. F. (2017). Variation in reproductive isolation across a
554 species range. *Ecology and Evolution*, *7*(22), 9347-9357.

555 Barney, J. N. (2006). North American History of Two Invasive Plant Species: Phytogeographic
556 Distribution, Dispersal Vectors, and Multiple Introductions. *Biological Invasions*, *8*(4), 703–
557 717.

558 Barrett, S. C. H., Colautti, R. I., & Eckert, C. G. (2008). Plant reproductive systems and evolution
559 during biological invasion. *Molecular Ecology*, *17*(1), 373–383.

560 Blackburn, T. M., Lockwood, J. L., & Cassey, P. (2015). The influence of numbers on invasion
561 success. *Molecular Ecology*, *24*(9), 1942–1953.

562 Blischak, P., Kubatko, L., & Wolfe, A. (2017). SNP genotyping and parameter estimation in
563 polyploids using low-coverage sequencing data. *Bioinformatics*, *34*(3), 407–415.

564 Bock, D. G., Kane, N. C., Ebert, D. P., & Rieseberg, L. H. (2014). Genome skimming reveals the
565 origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an
566 artichoke. *New Phytologist*, *201*(3), 1021–1030.

567 Burrell, A. M., Pepper, A. E., Hodnett, G., Goolsby, J. A., Overholt, W. A., Racelis, A. E., …
568 Klein, P. E. (2015). Exploring origins, invasion history and genetic diversity of *Imperata*
569 *cylindrica* (L.) P. Beauv.(Cogongrass) in the United States using genotyping by sequencing.
570 *Molecular Ecology*, *24*(9), 2177–2193.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

571 Butchart, S. H. M., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J. P. W., Almond, R.
572 E. A., … Bruno, J. (2010). Global biodiversity: indicators of recent declines. *Science*,
573 1187512.

574 Butlin, R. (2002). The costs and benefits of sex: new insights from old asexual lineages. *Nature*
575 *Reviews Genetics*, *3*(4), 311–317.

576 Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks:
577 building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes,*
578 *Genetics*, *1*(3), 171–182.

579 Chown, S. L., Hodgins, K. A., & Griffin, P. C. (2016). Biological invasions, climate change, and
580 genomics. In Crop Breeding (pp. 59-114). Apple Academic Press.

581 Combs, M., Puckett, E. E., Richardson, J., Mims, D., & Munshi-South, J. (2018). Spatial
582 population genomics of the brown rat (Rattus norvegicus) in New York City. *Molecular*
583 *Ecology*, *27*(1), 83–98.

584 Cornille, A., Salcedo, A., Kryvokhyzha, D., Glémin, S., Holm, K., Wright, S. I., & Lascoux, M.
585 (2016). Genomic signature of successful colonization of Eurasia by the allopolyploid
586 shepherd's purse (*Capsella bursa-pastoris*). *Molecular Ecology*, *25*(2), 616–629.

587 Coughlan, J. M., Han, S., Stefanović, S., & Dickinson, T. A. (2017). Widespread generalist
588 clones are associated with range and niche expansion in allopolyploids of Pacific Northwest
589 Hawthorns (*Crataegus* L.). *Molecular ecology*, *26*(20), 5484-5499.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

590 Del Tredici, P. (2017). The introduction of Japanese knotweed, *Reynoutria japonica*, into North
591 America. *The Journal of the Torrey Botanical Society*, *144*(4), 406–416.

592 Dlugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: genetic variation,
593 adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, *17*(1), 431–
594 449.

595 Dlugosch, K. M., & Parker, I. M. (2007). Molecular and quantitative trait variation across the
596 native range of the invasive species *Hypericum canariense*: evidence for ancient patterns of
597 colonization via pre-adaptation? *Molecular Ecology*, *16*(20), 4269–4283.

598 Dray, S., & Dufour, A.-B. (2007). The ade4 package: implementing the duality diagram for
599 ecologists. *Journal of Statistical Software*, *22*(4), 1–20.

600 Dufresne, F., Stift, M., Vergilino, R., & Mable, B. K. (2014). Recent progress and challenges in
601 population genetics of polyploid organisms: an overview of current state□of□the□art
602 molecular and statistical tools. *Molecular ecology*, *23*(1), 40-69.

603 Durka, W., Bossdorf, O., Prati, D., & Auge, H. (2005). Molecular evidence for multiple
604 introductions of garlic mustard (*Alliaria petiolata*, Brassicaceae) to North America.
605 *Molecular Ecology*, *14*(6), 1697–1706.

606 Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., ... &
607 Alger, E. I. (2019). Origin and evolution of the octoploid strawberry genome. *Nature*
608 *genetics*, *51*(3), 541.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

609 Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell,
610 S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity
611 species. *PloS One*, *6*(5), e19379.

612 Estoup, A., Ravigné, V., Hufbauer, R., Vitalis, R., Gautier, M., & Facon, B. (2016). Is there a
613 genetic paradox of biological invasion? *Annual Review of Ecology, Evolution, and*
614 *Systematics*, *47*, 51–72.

615 Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic
616 revolutions and structuration. *Trends in Ecology & Evolution*, *23*(7), 347–351.

617 Fisher, R. A. (1958). *The genetical theory of natural selection*. Oxford University Press.

618 Forman, J., & Kesseli, R. V. (2003). Sexual reproduction in the invasive species *Fallopia*
619 *japonica* (Polygonaceae). *American Journal of Botany*, *90*(4), 586–592.

620 Frankham, R. (1995). Inbreeding and extinction: a threshold effect. *Conservation Biology*, *9*(4),
621 792–799.

622 Frichot, E., & François, O. (2015). LEA: an R package for landscape and ecological association
623 studies. *Methods in Ecology and Evolution*, *6*(8), 925–929.

624 Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations
625 between loci and environmental gradients using latent factor mixed models. *Molecular*
626 *Biology and Evolution*, *30*(7), 1687–1699.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

627 Gammon, M. A., & Kesseli, R. (2010). Haplotypes of Fallopia introduced into the US. *Biological*
628 *invasions*, *12*(3), 421-427.

629 Gammon, M. A., Baack, E., Orth, J. F., & Kesseli, R. (2010). Viability, growth, and fertility of
630 knotweed cytotypes in North America. *Invasive Plant Science and Management*, *3*(3), 208–
631 218.

632 Gammon, M. A., Grimsby, J. L., Tsirelson, D., & Kesseli, R. (2007). Molecular and
633 morphological evidence reveals introgression in swarms of the invasive taxa *Fallopia*
634 *japonica*, *F. sachalinensis*, and *F.* x *bohemica* (Polygonaceae) in the United States.
635 *American Journal of Botany*, *94*(6), 948–56.

636 Gaskin, J. F., Schwarzländer, M., Grevstad, F. S., Haverhals, M. A., Bourchier, R. S., & Miller,
637 T. W. (2014). Extreme differences in population structure and genetic diversity for three
638 invasive congeners: knotweeds in western North America. *Biological Invasions*, *16*(10),
639 2127–2136.

640 Geng, Y.-P., Pan, X.-Y., Xu, C.-Y., Zhang, W.-J., Li, B., Chen, J.-K., … Song, Z.-P. (2007).
641 Phenotypic plasticity rather than locally adapted ecotypes allows the invasive alligator weed
642 to colonize a wide range of habitats. *Biological Invasions*, *9*(3), 245–256.

643 Grimsby, J., & Kesseli, R. (2010). Genetic composition of invasive Japanese knotweed sl in the
644 United States. *Biological Invasions*, *12*(7), 1943–1946.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

645 Grimsby, J. L., Tsirelson, D., Gammon, M. A., & Kesseli, R. (2007). Genetic diversity and clonal
646 vs. sexual reproduction in *Fallopia* spp. (Polygonaceae). *American Journal of Botany*, *94*(6),
647 957–64.

648 Gutekunst, J., Andriantsoa, R., Falckenhayn, C., Hanna, K., Stein, W., Rasamy, J., & Lyko, F.
649 (2018). Clonal genome evolution and rapid invasive spread of the marbled crayfish. *Nature*
650 *Ecology & Evolution*, *2*(3), 567.

651 Herman, J. J., & Sultan, S. E. (2011). Adaptive transgenerational plasticity in plants: case studies,
652 mechanisms, and implications for natural populations. *Frontiers in Plant Science*, *2*, 102.

653 Hollingsworth, M. L., & Bailey, J. P. (2000). Evidence for massive clonal growth in the invasive
654 weed *Fallopia japonica* (Japanese Knotweed). *Botanical Journal of the Linnean Society*,
655 *133*(4), 463–472.

656 Inamura, A., Ohashi, Y., Sato, E., Yoda, Y., Masuzawa, T., Ito, M., & Yoshinaga, K. (2000).
657 Intraspecific sequence variation of chloroplast DNA reflecting variety and geographical
658 distribution of *Polygonum cuspidatum* (Polygonaceae) in Japan. *Journal of Plant Research*,
659 *113*(4), 419–426.

660 Indyk, P., & Motwani, R. (1998). Approximate nearest neighbors: towards removing the curse of
661 dimensionality. In *Proceedings of the thirtieth annual ACM symposium on Theory of*
662 *computing* (pp. 604–613). ACM.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

663 Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers.
664 *Bioinformatics*, *24*(11), 1403–1405.

665 Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-
666 wide population genetic data with emphasis on clonality. *Frontiers in genetics*, *6*, 208.

667 Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12),
668 1225–1241.

669 Keller, S. R., & Taylor, D. R. (2010). Genomic admixture increases fitness during a biological
670 invasion. *Journal of Evolutionary Biology*, *23*(8), 1720–1731.

671 Kim, J. Y., & Park, C.-W. (2000). Morphological and chromosomal variation in *Fallopia* section
672 *Reynoutria* (Polygonaceae) in Korea. *Brittonia*, *52*(1), 34–48.

673 Klopfstein, S., Currat, M., & Excoffier, L. (2005). The fate of mutations surfing on the wave of a
674 range expansion. *Molecular Biology and Evolution*, *23*(3), 482–490.

675 Kolbe, J. J., Glor, R. E., Schettino, L. R., Lara, A. C., Larson, A., & Losos, J. B. (2004). Genetic
676 variation increases during biological invasion by a Cuban lizard. *Nature*, *431*(7005), 177–
677 181.

678 Lee, C. E. (2002). Evolutionary genetics of invasive species. *Trends in Ecology & Evolution*,
679 *17*(8), 386–391.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

680 Legendre, P., & Legendre, L. F. J. (2012). *Numerical ecology* (3rd ed., Vol. 24). Oxford, UK:
681 Elsevier.

682 Lowe, S., Browne, M., Boudjelas, S., & De Poorter, M. (2000). *100 of the world's worst invasive*
683 *alien species: a selection from the global invasive species database* (Vol. 12). Invasive
684 Species Specialist Group Auckland.

685 Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: an R package to perform genome scans for
686 selection based on principal component analysis. *Molecular ecology resources*, *17*(1), 67-
687 77.

688 Lynch, M. (1984). Destabilizing hybridization, general-purpose genotypes and geographic
689 parthenogenesis. *The Quarterly Review of Biology*, *59*(3), 257–290.

690 Mallet, J. (2007). Hybrid speciation. *Nature*, *446*(7133), 279.

691 Mayr, E. (1942). Systematics and the origin of species. Columbia Univ. Press, New York.
692 *Systematics and the Origin of Species. Columbia Univ. Press, New York.*

693 Meirmans, P. G., & van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: two programs
694 for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes, 4*(4),
695 792-794.

696 Meirmans, P. G., Liu, S., & van Tienderen, P. H. (2018). The analysis of polyploid genetic data.
697 *Journal of Heredity*, *109*(3), 283-296.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

698 Muirhead, C. A., & Lande, R. (1997). Inbreeding depression under joint selfing, outcrossing, and
699 asexuality. *Evolution*, *51*(5), 1409–1415.

700 Neiman, M., Sharbel, T. F., & Schwander, T. (2014). Genetic causes of transitions from sexual
701 reproduction to asexuality in plants and animals. *Journal of Evolutionary Biology*, *27*(7),
702 1346–1359.

703 Novak, S. J., & Mack, R. N. (2005). Genetic bottlenecks in alien plant species: influence of
704 mating systems and introduction dynamics. *Species Invasions: Insights into Ecology,*
705 *Evolution, and Biogeography*, 201–228.

706 Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H.
707 (2017). vegan: Community Ecology Package. R package version 2.4-4.

708 Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., &
709 Phillippy, A. M. (2016). Mash: fast genome and metagenome distance estimation using
710 MinHash. *Genome Biology*, *17*.

711 Otto, S. P., & Whitton, J. (2000). Polyploid incidence and evolution. *Annual Review of Genetics*,
712 *34*(1), 401–437.

713 Pandit, M. K., Tan, H. T. W., & Bisht, M. S. (2006). Polyploidy in invasive plant species of
714 Singapore. *Botanical Journal of the Linnean Society*, *151*(3), 395–403.

715 Pannell, J. R., & Barrett, S. C. H. (1998). Baker's law revisited: reproductive assurance in a
716 metapopulation. *Evolution*, *52*(3), 657–668.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

717 Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: a road map for
718 stacks. *Methods in Ecology and Evolution*, *8*(10), 1360–1373.

719 Peischl, S., Dupanloup, I., Kirkpatrick, M., & Excoffier, L. (2013). On the accumulation of
720 deleterious mutations during range expansions. *Molecular Ecology*, *22*(24), 5972–5982.

721 Prentis, P. J., Wilson, J. R. U., Dormontt, E. E., Richardson, D. M., & Lowe, A. J. (2008).
722 Adaptive evolution in invasive species. *Trends in Plant Science*, *13*(6), 288–294.

723 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using
724 multilocus genotype data. *Genetics*, *155*(2), 945–959.

725 Razanajatovo, M., & van Kleunen, M. (2016). Non-invasive naturalized alien plants were not
726 more pollen-limited than invasive aliens and natives in a common garden. *Functional*
727 *Ecology*, *30*(9), 1511–1520.

728 Reusch, T. B. H., Stam, W. T., & Olsen, J. L. (2000). A microsatellite-based estimation of clonal
729 diversity and population subdivision in *Zostera marina*, a marine flowering plant. *Molecular*
730 *Ecology*, *9*(2), 127–140.

731 Richards, C. L., Schrey, A. W., & Pigliucci, M. (2012). Invasion of diverse habitats by few
732 Japanese knotweed genotypes is correlated with epigenetic differentiation. *Ecology Letters*,
733 *15*(9), 1016–1025.

734 Richards, C. L., Walls, R. L., Bailey, J. P., Parameswaran, R., George, T., & Pigliucci, M.
735 (2008). Plasticity in salt tolerance traits allows for invasion of novel habitat by Japanese

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

736 knotweed s. l. (*Fallopia japonica* and *F.* x *bohemica*, Polygonaceae). *American Journal of*
737 *Botany*, *95*(8), 931–42.

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

740 Roberts, A. V. (2007). The use of bead beating to prepare suspensions of nuclei for flow
741 cytometry from fresh leaves, herbarium leaves, petals and pollen. *Cytometry Part A*, *71*(12),
742 1039–1044.

743 Roman, J., & Darling, J. A. (2007). Paradox lost: genetic diversity and the success of aquatic
744 invasions. *Trends in Ecology & Evolution*, *22*(9), 454–464.

745 Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under
746 isolation by distance. *Genetics*, *145*(4), 1219–1228.

747 Schliep, K. P. (2010). phangorn: phylogenetic analysis in R. *Bioinformatics*, *27*(4), 592-593.

748 Schoen, D. J., & Schultz, S. T. (2019). Somatic mutation and evolution in plants. *Annual Review*
749 *of Ecology, Evolution, and Systematics*, *50*, 49-73.

750 Sim, S. C., Kim, M., Chung, S. M., & Park, Y. (2015). Assessing the genetic variation in
751 cultivated tomatoes (*Solanum lycopersicum* L.) using genome-wide single nucleotide
752 polymorphisms. *Horticulture, Environment, and Biotechnology*, *56*(6), 800-810.

753 Soltis, P. S., & Soltis, D. E. (2000). The role of genetic and genomic attributes in the success of
754 polyploids. *Proceedings of the National Academy of Sciences*, *97*(13), 7051–7057.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

755 Soltis, P. S., & Soltis, D. E. (2009). The role of hybridization in plant speciation. *Annual Review*
756 *of Plant Biology*, *60*, 561–588.

757 Stouthamer, R., Breeuwer, J. A. J., & Hurst, G. D. D. (1999). Wolbachia pipientis: microbial
758 manipulator of arthropod reproduction. *Annual Reviews in Microbiology*, *53*(1), 71–102.

759 te Beest, M., Le Roux, J. J., Richardson, D. M., Brysting, A. K., Suda, J., Kubešová, M., &
760 Pyšek, P. (2011). The more the better? The role of polyploidy in facilitating plant invasions.
761 *Annals of Botany*, *109*(1), 19–45.

762 Thompson, J. D., McNeilly, T., & Gray, A. J. (1991). Population variation in *Spartina anglica*
763 CE Hubbard. *New Phytologist*, *117*(1), 141–152.

764 Trumbo, D. R., Epstein, B., Hohenlohe, P. A., Alford, R. A., Schwarzkopf, L., & Storfer, A.
765 (2016). Mixed population genomics support for the central marginal hypothesis across the
766 invasive range of the cane toad (*Rhinella marina*) in Australia. *Molecular Ecology*, *25*(17),
767 4161–4176.

768 Uller, T., & Leimu, R. (2011). Founder events predict changes in genetic diversity during
769 human-mediated range expansions. *Global Change Biology*, *17*(11), 3478–3485.

770 VanWallendael, A., Alvarez, M., & Franks, S.F. (2019). Japanese knotweed GBS. NCBI SRA
771 BioProject PRJNA574173

772 Vekemans, X., & Hardy, O. J. (2004). New insights from fine-scale spatial genetic structure
773 analyses in plant populations. *Molecular Ecology*, *13*(4), 921–935.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

774 Verhoeven, K. J. F., Macel, M., Wolfe, L. M., & Biere, A. (2011). Population admixture,
775 biological invasions and the balance between local adaptation and inbreeding depression.
776 *Proceedings of the Royal Society of London B: Biological Sciences*, *278*(1702), 2–8.

777 Vilà, M., Weber, E., & Antonio, C. M. D. (2000). Conservation implications of invasion by plant
778 hybridization. *Biological Invasions*, *2*(3), 207–217.

779 Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan
780 methods against more complex models: when and how much should we trust them?
781 *Molecular Ecology*, *23*(8), 2006–2019.

782 Voichek, Y. & Weigel, D., (2019). Finding genetic variants in plants without complete
783 genomes. *bioRxiv*, p.818096. https://doi.org/10.1101/818096.

784 Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some
785 genetic methods for identifying the number of gene pools and their degree of connectivity.
786 *Molecular Ecology*, *15*(6), 1419.

787 Whitham, T. G., & Slobodchikoff, C. N. (1981). Evolution by individuals, plant-herbivore
788 interactions, and mosaics of genetic variability: the adaptive significance of somatic
789 mutations in plants. *Oecologia*, *49*(3), 287-292.

790 Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97–159.

791 Zika, P. F., & Jacobson, A. L. (2003). An overlooked hybrid Japanese knotweed (Polygonum
792 cuspidatum× sachalinense; Polygonaceae) in North America. *Rhodora*, 143–152.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

793

794 **Tables and Figures**

795

796 **Table 1.** STACKS-derived diversity data for individual species. H_e: Expected Heterozygosity; π:

797 Nucleotide diversity; SE: Standard Error

798

Species	Ploidy	# Samples	Total Loci	Polymorphic Loci	% Polymorphic Loci
Japanese	8	55	1.95×10^7	12980	0.665
Bohemian	6	18	1.10×10^7	11711	1.061
Giant	4	15	2.05×10^7	21540	1.046

Species	Private SNPs	H_e	SE	π variant	π invariant
Japanese	1091	0.279	0.002	0.283 ± 0.002	0.0024 ± 0.0001
Bohemian	858	0.303	0.001	0.318 ± 0.001	0.0035 ± 0.0001
Giant	1617	0.309	0.001	0.326 ± 0.001	0.0038 ± 0.0001

799

800

801

802

803

804

805

806

807

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

808 **Table 2.** PERMANOVA results for two genetic distance estimation methods, SNP-based Nei's
809 distance, and *Mash* distance from the *Mash* MinHash algorithm. Permutational model selection
810 determined the best model to explain the Nei's distance matrix as Species + Site + Habitat, and
811 the *Mash* distance matrix as Species + Site + Sex. *F*- and *p*- values shown are marginal values
812 when each factor is tested against all other factors. Bold values were significant at $\alpha = 0.05$,
813 italicized values are marginally significant at $\alpha = 0.1$
814

Nei distance			***Mash* distance**		
Factor	***F***	***p***	**Factor**	***F***	***p***
Species	**7.31**	**0.001**	**Species**	**2.83**	**0.001**
Site	**1.56**	**0.009**	**Site**	**1.19**	**0.001**
Habitat	*2.29*	*0.064*	*Sex*	*1.30*	*0.071*
Overall	**3.38**	**0.001**	**Overall**	**1.87**	**0.001**
Sex	0.729	0.715	Habitat	1.08	0.263
Distance	0.907	0.679	Distance	1.03	0.387

815
816
817
818
819
820
821
822
823

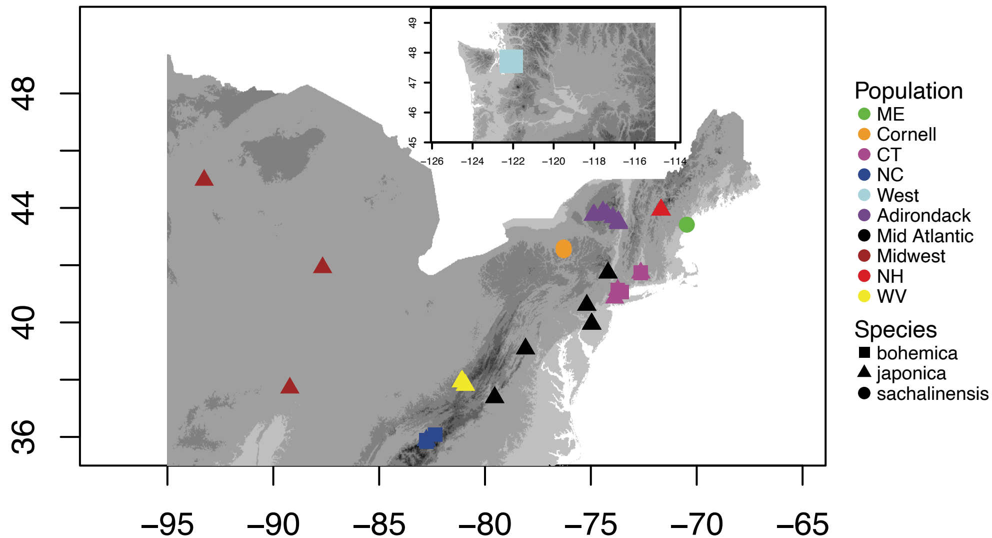
bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.226886; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

824 **Figure Legends**

825 **Figure 1**. Sample collection locations in eastern North America. Population identity is shown by
826 color, species is shown by shape. Sites CT and NC contained both *R.* x *bohemica* and *R.*
827 *sachalinensis*. Groups used in analyses are shown by color. Inset shows the single West coast
828 sample, collected from Seattle, WA.
829
830

831 **Figure 2.** Principal Coordinates Analysis of distance matrices produced by two methods.
832 Projected inertia is shown with the axis labels. Species identification was based on morphological
833 differences. (a) Nei's distance matrix, produced from SNP calls in STACKS (b) Distance matrix
834 produced by MinHash distance algorithm in *Mash*. (c) *Reynoutria japonica* (8X) analyzed alone,
835 using Nei's distance. (d) *Reynoutria japonica* (8X) analyzed alone, using *Mash* distance.
836

837 **Figure 3**. Maximum likelihood tree and ADMIXTURE results. (a): ML tree from SNP calls. *R.*
838 *sachalinensis* is shown in red, *R. japonica* in green, and *R.* x *bohemica* in blue. Bootstrap values
839 higher than 50 are shown at nodes. (b): Admixture assuming K = 2, 3, and 4 ancestral
840 populations. Each bar represents one sample. (c) Coefficient of variation (CV) for each tested
841 value of K for admixture analysis. Dashed lines show K = 2-4. (d) Geographic and Nei's genetic
842 distance (D) correlation.
843

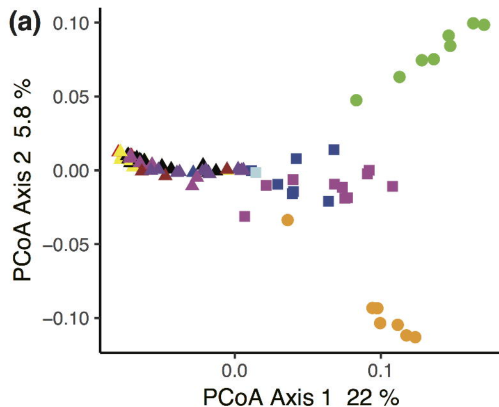


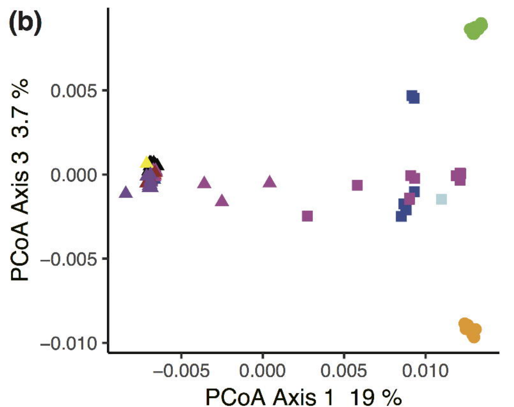
48
44
40
36
−95
−90
−85
−80
−75
−70
−65
49
48
47
46
45
−126
−124
−122
−120
−118
−116
−114
Population
ME
Cornell
CT
NC
West
Adirondack
Mid Atlantic
Midwest
NH
WV
Species
bohemica
japonica
sachalinensis

Nei's Distance

Mash Distance

All samples

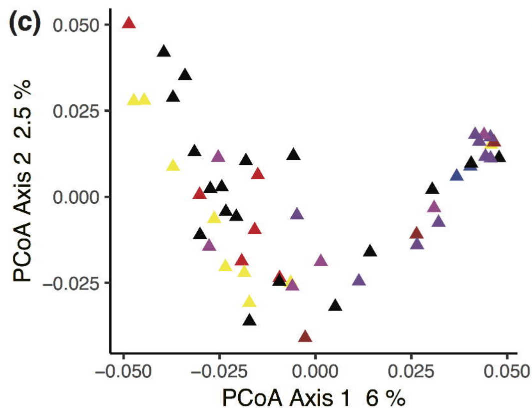


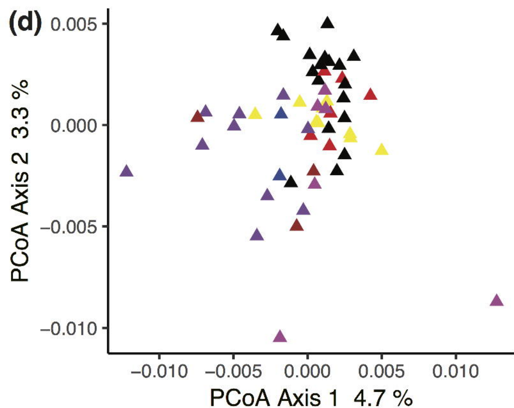


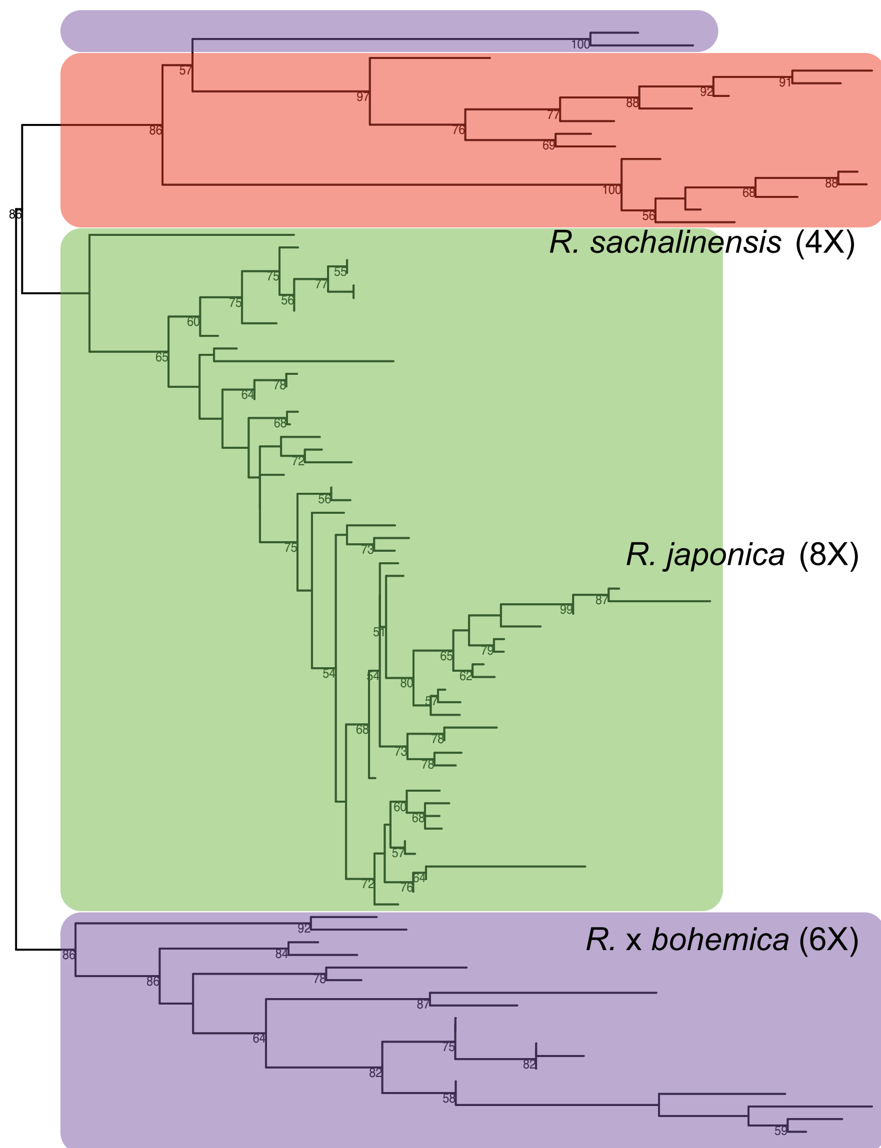
Population
- ME
- Cornell
- CT
- NC
- West
- Adirondack
- Mid Atlantic
- Midwest
- NH
- WV

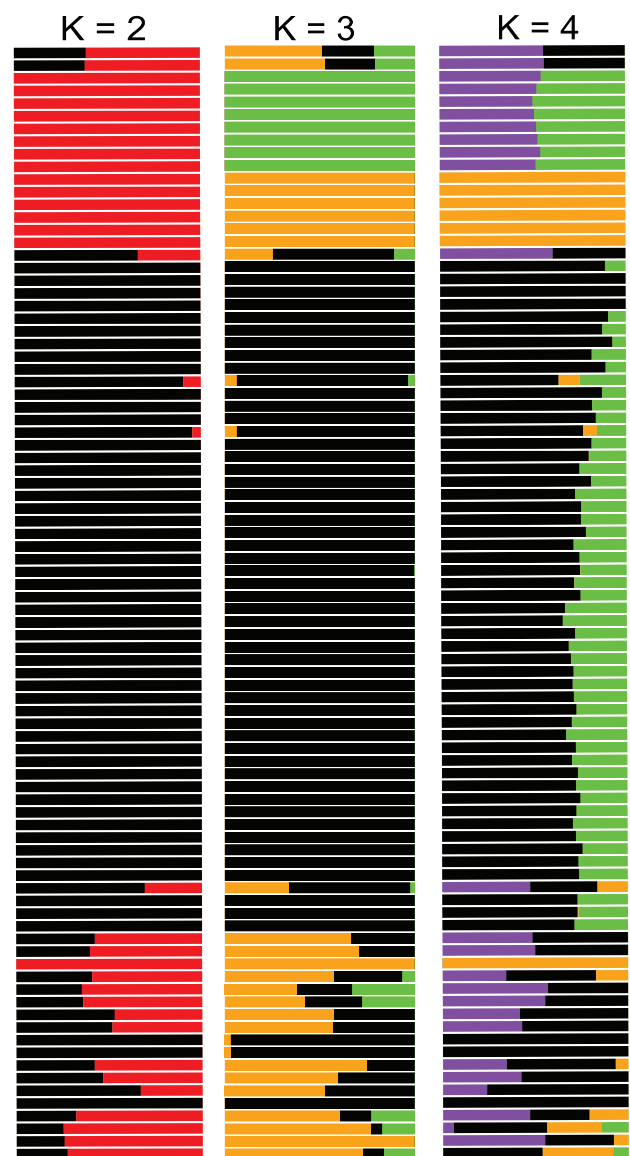
Species
- bohemica
- japonica
- sachalinensis

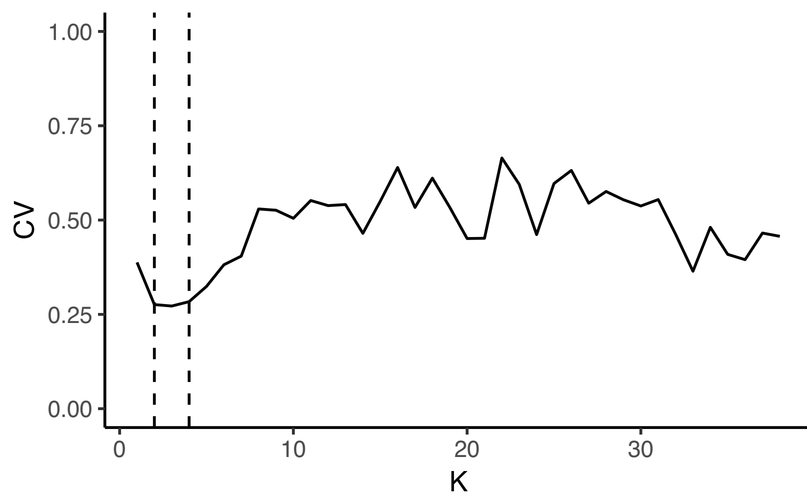
R. japonica only

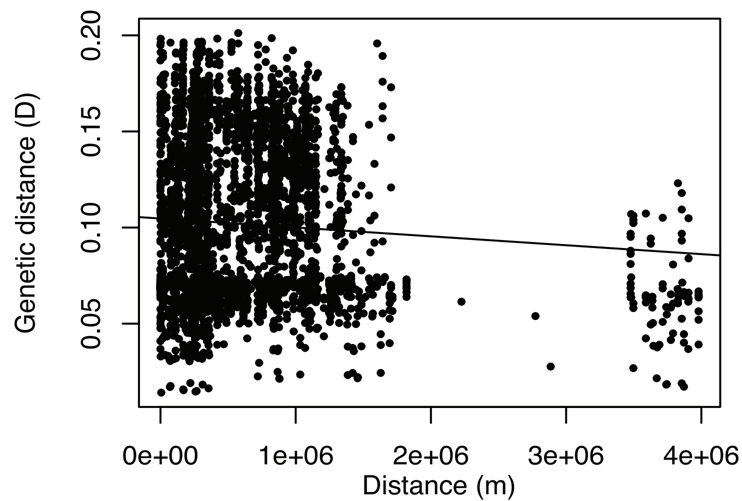





(a)
R. sachalinensis (4X)
R. japonica (8X)
R. x bohemica (6X)


(b)
K = 2
K = 3
K = 4


(c)
CV
K
1.00
0.75
0.50
0.25
0.00
0
10
20
30


(d)
Genetic distance (D)
Distance (m)
0.20
0.15
0.10
0.05
0e+00
1e+06
2e+06
3e+06
4e+06