The tip of the iceberg: genome wide marker analysis reveals hidden hybridization during invasion

Running title: Hybridization of co-invaders

Hanna S. Rosinger[1], Armando M. Geraldes[2,4], Kristin A. Nurkowski[1,4], Paul Battlay[1], Roger D. Cousens[3], Loren H. Rieseberg[4], Kathryn A. Hodgins[1]

[1] School of Biological Sciences, Monash University, Melbourne, VIC, Australia

[2]Department Zoology, University of British Columbia, Vancouver, BC, Canada,

[3]School of BioSciences, University of Melbourne, Melbourne, VIC, Australia

[4] Department of Botany and Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada

*** Correspondence:**

Hanna S. Rosinger

Hanna.Rosinger@monash.edu

Abstract

Biological invasions are accelerating, and invasive species can have large economic impacts as well as severe consequences for biodiversity. During invasions, species can interact, potentially resulting in hybridization. Here, we examined two *Cakile* species, *C. edentula* and *C. maritima* (Brassicaceae), that co-occur and may hybridize during range expansion in separate regions of the globe. *Cakile edentula* invaded each location first, while *C. maritima* established later, apparently replacing the former. We assessed the evidence for hybridization in western North America and Australia, where both species have been introduced, and identified source populations with 4561 SNPs using Genotype-by-Sequencing. Our results indicate that the *C. edentula* in Australia originated from one region of eastern North America while in western North America it is likely from multiple sources. The *C. maritima* in Australia were derived from at least two different parts of Europe while the introduction in western North America is from one. Although morphological evidence of hybridization is generally limited to mixed species populations in Australia and virtually absent elsewhere, our genetic analysis revealed relatively high levels of hybridization in Australia (58% hybrids) and supported the presence of hybrids in western North America (16%) and New Zealand. Hybrids might be commonly overlooked in invaders, as identification based solely on morphological traits may represent only the tip of the iceberg. Our study reveals a repeated pattern of invasion, hybridization and apparent replacement of one species by another, which offers an opportunity to investigate the role of hybridization and introgression during invasion.

Keywords: invasion, hybridization, *Cakile edentula*, *Cakile maritima*, Genotype-by-Sequencing (GBS), range expansion

1 Introduction

Biogeographic barriers on a global, regional and local scale are often overcome by human activities,

leading to biological invasions (Sax & Gaines, 2003; Simberloff, 2013; Vilatersana, Sanz, Galian, & Castells, 2016). Biological invasions can have a large economic impact, reaching into the billions (Hoffmann & Broadhurst, 2016; Pimentel, Zuniga, & Morrison, 2005), as well as severe negative consequences for biodiversity and ecosystems (Sakai et al., 2001). Most long-distance introductions of invasive species in historic times are directly (e.g. ornamentals) or indirectly the result of anthropogenic activities (e.g. ballast on ships) (Baker, 1974; Ruiz et al., 2000; Sakai et al., 2001). Invasions can also lead to novel interactions between species that previously had not co-occurred and, where there are no strong reproductive barriers, this may lead to instances of hybridization (Abbott, 1992; Ellstrand & Schierenbeck, 2000).

Rather than hybridization just being an incidental event, it could actually facilitate the success of invasive plant species, as invasive hybrid lineages can have increased fecundity and size (Hovick & Whitney, 2014). Various hypotheses have been proposed by which hybridization facilitates rapid range expansion (Bock et al., 2015; Ellstrand & Schierenbeck, 2000), including evolutionary novelty, increased genetic variation, heterosis, dumping genetic load (i.e. genetic rescue) (Ellstrand & Schierenbeck, 2000) and demographic rescue (Mesgaran et al., 2016). But convincing empirical data are limited. Hybridization is certainly not the sole evolutionary pathway to invasiveness, but it can catalyze the evolution of invasiveness (Ellstrand & Schierenbeck, 2000). Not all the potential consequences of hybridization are beneficial, however, and there can be significant costs associated with hybridization, such as outbreeding depression (Baack, Melo, Rieseberg, & Ortiz‐Barrientos, 2015) and genetic swamping (Todesco et al., 2016). Our capacity to assess the role of hybridization during any particular invasion is hampered by the fact that it can be difficult to identify, especially when repeated backcrossing with one parental species has occurred rendering morphological identification difficult (Ward, Gaskin, & Wilson, 2008). However, genome-wide molecular markers can provide estimates of the extent of past hybridization and introgression across the genome

(Payseur & Rieseberg, 2016).

On the beaches of Australia, the North Island of New Zealand and western North America a repeated pattern of invasion by two species of sea-rocket with contrasting mating systems (Barbour & Rodman, 1970; Cousens, Ades, Mesgaran, & Ohadi, 2013; Cousens & Cousens, 2011; Rodman, 1974, 1986) offers a rare opportunity to investigate the role of hybridization during invasion in distinct, geographically isolated regions. *Cakile edentula* (American sea-rocket), native to eastern North America, invaded each location first, while *Cakile maritima* (European sea-rocket) (Brassicaceae), native to Europe and northern Africa, arrived later. The invasion and replacement history in western North America and Australia are reviewed elsewhere (Barbour & Rodman, 1970; Cousens et al., 2013; Rodman, 1986), but we briefly outline it below.

In Australia, *C. edentula* was first recorded in Victoria in 1863 and subsequently spread along the coastline of Australia (Rodman, 1986). In 1897, *C. maritima* was recorded for the first time in Western Australia, and a second introduction into South Australia (1918: see Cousens et al., 2013; Ohadi et al., 2016) spread from there to the east (Heyligers, 1984; Rodman, 1986). In contrast to *C. edentula*, *C. maritima* seems still to be actively spreading in Australia and appears to have replaced *C. edentula* throughout much of its initial introduced range (Cousenes et al., 2013; Rodman, 1986). In western North America, a similar pattern of replacement occurred. *Cakile edentula* was found near San Francisco around 1880 (Barbour & Rodman, 1970), while *C. maritima* reached western North America by 1936 where it was found sympatric with *C. edentula* near San Francisco. The most recent published field study showed that *C. maritima* had replaced *C. edentula* throughout most of coastal California but not Oregon or Washington (Boyd & Barbour, 1993). In each case, there has been complete replacement of *C. edentula* by *C. maritima* over wide geographic areas (Barbour & Rodman, 1970; Cousens et al., 2013; Rodman, 1986), which was originally assumed to involve either direct or indirect competition (Rodman, 1986), although several additional mechanisms have been

85 proposed such as disease (Bock, 2008; Cousens et al., 2013; Thrall, Young, & Burdon, 2000),
86 coincidence (Cousens et al., 2013; Rodman, 1986), or greater lifetime fecundity of *C. maritima*
87 (Boyd & Barbour, 1993). However, the mechanism of the replacement remains unclear.

88 *Cakile edentula* and *C. maritima* are closely related and cross-compatible (Li, Cousens, & Mesgaran,
89 2019; Mesgaran et al., 2016; Rodman, 1974). Both species are found in coastal strandline habitat,
90 providing opportunities for hybridization in regions where they co-occur, but the species exhibit
91 contrasting mating systems (Rodman, 1974). *Cakile edentula* (self-compatible) benefits from high
92 levels of reproductive assurance as it is able to set seeds autonomously at high rates (Li, Mesgaran,
93 Ades, & Cousens, 2020); one of Baker's (1965) ideal weed traits. In contrast, the establishment of *C.*
94 *maritima* (self-incompatible) may be initially hindered (during both initial establishment as well as
95 subsequent range expansion) by a lack of compatible mates limiting sexual reproduction and
96 resulting in strong Allee effects. The apparent presence of hybrids, based on an intermediate leaf and
97 fruit shape of both parental species, in some sites in Australia led Mesgaran et al., (2016) to develop a
98 model for the interacting species, with the novel outcome that transient hybridization could overcome
99 Allee effects in *C. maritima*. As a consequence, we hypothesized that past hybridization with *C.*
100 *edentula* could be a common feature of *C. maritima's* establishment and range expansion in western
101 North America, Australia and New Zealand.

102 We used genome-wide markers derived from Genotype-by-Sequencing (GBS) to examine the
103 invasion history of these two species in Australia and western North America and quantify the extent
104 and distribution of hybridization. There have been several previous studies examining the population
105 genetic structure of *C. edentula* and *C. maritima* in their native ranges (Europe (Clausing, Vickers,
106 Kadereit, 2000; Kadereit, Arafeh, Somogyi, & Westberg, 2005; Westberg, 2005), Africa (Gandour,
107 Hessini, & Abdelly, 2008), eastern and western North America (Gormally, Hamrick, & Donovan,
108 2011) as well as in the introduced range of Australia (Ohadi et al., 2016). However, no study of the

invasion history on two continents has been attempted nor has the extent of hybridization across multiple introductions been quantified. Specifically, we aimed to (1) identify probable source regions (from Europe and eastern North America); (2) determine whether both recent and advanced generation hybrids occur in the introduced ranges and the extent of their geographic distribution; and (3) determine if the change in levels of species ancestry post-invasion reflects a chronosequence along the direction of invasion of *C. maritima*. We predicted that early generation hybrids should be present at the leading edge of *C. maritima's* invasion into *C. edentula*-occupied areas, but later generation backcrosses with *C. maritima* should be more common in areas closer to where *C. maritima* first established. This should contribute to a gradient in species ancestry whereby *C. maritima* ancestry will be dominant in hybrids near the invasion source, while *C. edentula* ancestry will be more prevalent in hybrids identified in areas recently invaded by *C. maritima*. We predicted high levels of *C. maritima* ancestry in hybrids near the invasion source because *C. maritima* phenotypes are now exclusively present in the regions surrounding the invasion source, and studies of pollinators suggest preferential visitation of both hybrids and *C. maritima* over *C. edentula* which should facilitate backcrossing to *C. maritima* (Mesgaran et al., 2016).

2 Methods

2.1 Study species

Cakile maritima's native range extends over a wide climatic range from northern Norway to northern Africa. Current taxonomy recognizes subsp. *maritima* in the Mediterranean, subsp. *baltica* in the Baltic, subsp. *integrifolia* on the Atlantic coast and subsp. *euxina* in the Black Sea (Marhold, 2011). This is paralleled in the western Atlantic by *C. edentula,* which is found from Labrador to northern Florida, and two subspecies are recognized in its native range (Rodman, 1974) subsp. *edentula* (Labrador to North Carolina) and subsp. *harperi* (North Carolina to Florida). Both species exhibit

variation in morphology that is structured geographically (Ball, 1964; Rodman, 1974). Although *C. maritima* has a sporophytic self-incompatibility system, the level of self-incompatibility varies among plants (Thrall et al., 2000). *Cakile edentula* is self-compatible and can set seed autonomously at a high rate (Barbour, 1970; Rodman, 1974), although field estimates are suggestive of intermediate levels of autonomous selfing (Li et al., 2020). Anthers of *C. edentula* dehisce before the flowers open indicating opportunities for prior selfing (Li et al., 2019). Both species are diploid (2n = 18) (Rodman, 1974). Hybrids are readily produced through artificial pollination (Rodman, 1974) and with either parent as the pollen donor when emasculated (Li et al., 2019; Mesgaran et al., 2016), although crosses are more successful when *C. edentula* acts as the pollen recipient, consistent with the SI x SC rule (Harrison & Darby, 1955).

2.2 Samples

Samples of *Cakile spp*. were obtained from the native ranges (Europe and northern Africa, eastern North America) and the two introduced ranges (Australasia, western North America). We collected four of the five subspecies (subsp. *baltica*, subsp. *maritima*, subsp. *integrifolia* and subsp. *islandica*) of *C. maritima* (exclusion of subsp. *euxina*). In the native range of *C. edentula* we sampled only *C. edentula* subsp. *edentula* as this subspecies is most likely the source of invasions in Australia and western North America (Cousens et al., 2013; Rodman, 1974). We obtained 214 samples of *C. maritima*, 137 samples of *C. edentula*, 17 putative hybrids (identified by morphology in the field) and two *C. lanceolata* samples. Samples were sourced from 92 locations in total (Figure S1; Table 1 & S1). Many of these samples were our own field collections of silica dried leaf tissue (particularly in the introduced ranges), although a few samples were purified DNA from colleagues. We collected our samples along a transect through a population, ensuring that individuals were at least 2 m apart to

avoid sampling close relatives or the same individual. Individuals were collected randomly with respect to their putative species based on morphology.

2.3 DNA extraction and Genotype-by-sequencing

We performed DNA extractions from dried leaf material using a modified CCDB DNA Extraction Protocol following Whitlock, Hipperson, Mannarelli, and Burke (2008). DNA quantity was assessed using a QuBit broad-sensitivity DNA quantification system (Invitrogen, Carlsbad, CA, USA) and a double-digest GBS library preparation was carried out (using PstI-HF (NEB) and MspI (NEB) enzymes, see Supplementary Information for details). Sequencing (125bp PE) was conducted on an Illumina HiSeq2500 (McGill University and Genome Quebec Innovation Centre) on two lanes.

2.4 SNP calling

Quality statistics of raw reads were assessed though FastQC (http://hannonlab.cshl.edu/fastx_toolkit) and the reads were demultiplexed using STACKS process_radtags (Catchen, Amores, Hohenlohe, Cresco, & Postlethwait, 2011). We removed adapter sequences and trimmed the reads using Sickle (Joshi & Fass, 2011) with a Q-score of ≥ 20 and read length of ≥ 20 base pair. FASTQ quality filter (http://hannonlab.cshl.edu/fastx_toolkit) was then used to filter for reads with a Q- score of 20 or greater for ≥ 90% of the read length. The filtered reads were aligned using the Burrows-Wheeler Aligner (BWA) (Li & Durbin, 2009) to a *C. maritima* draft genome. Early access to the draft genome was generously provided by S.I. Wright, University of Toronto (https://genome.jgi.doe.gov/portal/CakmarStandDraft/CakmarStandDraft.info.html, GenBank: MK637688.1). The current assembly of the reference genome is found in 26,153 scaffolds with a

scaffold N50 of 85,425. We assessed if there was a bias when mapping the reads of *C. edentula* to the reference genome of *C. maritima* but found limited evidence for such a bias (see Supplementary Information for further details).

We called variants with GATK HaplotypeCaller (Poplin et al., 2017). We refer to this as the *unfiltered dataset* (Rosinger et al., 2020). Using VCFtools (Danecek et al., 2011) we removed individuals with fewer than 25000 reads, removed indels and restricted individual genotypes to have a depth between 5- 100,000. Furthermore, we filtered for a minimum quality score of 20, a genotype quality of 20, and a minor allele frequency of 0.05. Subsequently, we kept only bi-allelic variants that were successfully genotyped in more than 50% of individuals and removed individuals that had more than 50% missing data. The above filtering steps resulted in a reduction from 699,585 SNPs in 371 individuals to 18,573 SNPs in 258 individuals. Additionally, we removed 121 SNPs which showed > 80% observed heterozygosity, because such high observed heterozygosity could be caused by paralogues. We refer to this as the *filtered dataset* (Rosinger et al., 2020), which had a mean coverage of 39.21 (minimum coverage 9.18, maximum coverage 504.73).

2.5 Genetic clustering

Population genetic structure was inferred using Admixture (Alexander, Novembre, & Lange, 2009). For Admixture and most of our analysis we thinned our *filtered dataset* for linkage using a single SNP per 1kb window, resulting in a reduction to 4561 SNPs from 257 individuals (excluding the outgroup *C. lanceolata*). We will refer to this as the *global thinned dataset*. We ran Admixture using the *global thinned dataset* with a major termination criterion of $1X10^{-9}$ (i.e., a run was terminated

199 when the change in log-likelihood between successive iterations was below $1X10^{-9}$), 1000 bootstraps
200 and ten-fold cross-validation for K=1-10. The K that produced the lowest cross-validation error was
201 selected as the best K value. We refer to this as the *unsupervised run*. All following analyses were
202 conducted in R-studio v.1.1.414 (RStudio Team, 2015) except where otherwise stated. The output of
203 Admixture visualized with pophelper v.2.3.0 (Francis, 2017) and pie charts.

204 To complement the population clustering analysis provided by Admixture, and provide further
205 insight in the population differentiation, we conducted a principal component analysis (PCA) and an
206 unrooted phylogenetic network analysis. Genetic differentiation between native and introduced
207 populations was summarized in a PCA using the R package SNPRelate (Zheng et al., 2012) and
208 tidyverse (Wickham, Francois, Henry, & Müller, 2019). The 95% confidence ellipse construction
209 was carried out using the R package car (Fox & Weisberg, 2019). We conducted this analysis using
210 the *global thinned dataset*. We used SPLITSTREE5 (Huson & Bryant, 2006) to visualize the overall
211 sample relatedness with an unrooted phylogenetic network. To do this, we created two datasets from
212 our unfiltered dataset; (1) a global dataset containing all samples (*global Splitstree dataset*) and (2) a
213 native range dataset containing samples from Europe and eastern North America (*native range*
214 *Splitstree dataset*). The above two datasets were created by filtering the *unfiltered dataset* for a minor
215 allele count of 2, a minimum genotype quality of 20 and a maximum missing value of 1. This
216 approach kept variants specific to the *C. lanceolata* lineage, which would have been removed by the
217 previous filtering steps. VCFtools (Danecek et al., 2011) and Mesquite (Maddison & Maddison,
218 2019) were used for filtering and data conversion.

219

220 **2.6 Hybrid identification**

221 We used three different approaches to identify hybrids using genetic data:

(1) To identify the proportion of each individual's genome that was attributable to each species' ancestry, we conducted a *supervised run* of Admixture for K=2 using the *global thinned dataset*, by setting the samples from the two native ranges as reference individuals. Providing known ancestries allows the program to set some rows in the matrix Q to known constants and provides a more accurate estimation of the ancestries of the remaining individuals, and of the ancestral allele frequencies (Alexander et al., 2009). These reference individuals are essentially training samples, and ancestry identification is transformed into a supervised learning problem. The other settings were retained from the *unsupervised run*. We refer to this as the *supervised run* and used this run to classify individuals by their Q-scores as hybrid, or pure species. We used the highest standard error from the Q scores obtained with 1000 bootstraps, resulting in individuals classified as hybrids if $0.025 < Q > 0.975$ of their genome was assigned to the *C. edentula* cluster.

(2) We used the program NewHybrids (Anderson & Thompson, 2002) to identify early generation hybrids as we expected early generation hybrids to be present in mixed species populations. It classifies their generation using a Bayesian model-based clustering framework to compute, by Markov chain Monte Carlo, the posterior probability that each individual belongs to each of the distinct hybrid classes (parental species, F1, F2, BC to species 1, BC to species 2). This program is designed to identify hybrids from the first two generations of interbreeding based on classification into six genotype classes and does not require the loci to be fixed between species, although a large number of highly differentiated loci aids hybrid identification (Anderson & Thompson, 2002). As the program is unable to deal with a large dataset, we restricted our data to 63 SNPs that showed fixed differences between the two species obtained from individuals classified as parental species using the *supervised run* of Admixture. Details of the settings used are provided in the Supplementary Information.

(3) We used the R package HIest (Fitzpatrick, 2012), which uses maximum likelihood to estimate

ancestry and heterozygosity. This method jointly considers ancestry (similar to Admixture) together with interclass heterozygosity (proportion of loci with alleles from both ancestral populations) and without the assumption that only two generations of admixture have transpired. It specifically tests the assumption that discrete classification (i.e., pure species or early generation hybrids) rather than continuous distribution of hybrid genotypes best describes each individual. The simple likelihood approach it employs is relatively robust to small errors in the assumed parental allele frequencies, especially if the errors are unbiased. For this package, we used the 471 loci that showed fixed differences between the individuals of the native ranges. Because it is possible that there is a low level of segregating variation within each species for these loci due to sampling error, particularly for SI *C. maritima* where the sample size is lower, we set the allele frequencies as 0.99 for *C. edentula* and 0.03 for *C maritima*. We also tested other SNP sets and allele frequencies. The details of the settings used and the hybrid assignments are provided in the Supplementary Information, but the patterns were broadly similar among runs.

We tested for a chronosequence by assessing if there was a correlation between the distance of each population from the first entry point of *C. maritima* (Adelaide in Australia, San Francisco in western North America) and the level of *C. maritima* and *C. edentula* ancestry using a Spearman's rank correlation test in R using the ggpubr package (Kassambara, 2020). We used the ranked order of populations from this origin point along the coastline for each range. In Australia, we only used the south-east mainland individuals, as the introduction history and pattern of replacement based on herbarium records led us to predict a gradient in species ancestry in hybrids from high levels of *C. maritima* in South Australia to high levels of *C. edentula* further north in Queensland. In western North America we predicted this pattern to the north of San Francisco (the likely origin of *C. maritima*) as *C. edentula* has only recently been replaced in parts of Oregon and Washington and *C. edentula* is common in British Columbia. We tested the correlation between the Q value of the *C.*

edentula cluster of the *supervised run* for each population and the rank order of the sampling locations along the coastline to the first entry point of *C. maritima*. We used individuals that were classified as hybrids by Admixture or all samples (including the parental species). We repeated this analysis using the S value from HIest and the hybrid classifications of this program.

Additionally, we used the program TreeMix (Pickrell & Pritchard, 2012) to identify evidence for hybridization in the introduced ranges using the *global thinned dataset*. We constructed maximum likelihood trees with TreeMix (Pickrell & Pritchard, 2012) allowing up to four migration events. First, we grouped our samples according to their species and origin. For *C. maritima*, we kept the Atlantic and Mediterranean *C. maritima* samples separate because they were likely different subspecies (Rodman, 1974, 1976, 1986) and these groups appeared well differentiated from one another (e.g., Figure 2 B). We excluded morphological hybrids, which appeared to be mainly early generation hybrids based on the NewHybrid analysis, to assess evidence for admixture between the species in the introduced ranges, which may not be apparent phenotypically. Our groupings were: 1) Australian *C. edentula*; 2) Australian *C. maritima* (Mediterranean); 3) Australian *C. maritima* (Atlantic); 4) western North American *C. edentula*, 5) western North American *C. maritima*; 6) eastern North American *C. edentula*; 7) European *C. maritima* (Mediterranean); and 8) European *C. maritima* (Atlantic). We tested for admixture in Australia separately from western North America but included native range samples in both analyses. We used the f_3 statistic (Pickrell & Pritchard, 2012; Reich, Thangaraj, Patterson, Price, & Singh, 2009), which is part of the TreeMix package, to test for evidence of admixture in the invasive ranges in putative hybrids. We grouped the samples according to their Admixture classification (*supervised run*). For Australia we had three groups in south-east Australia: 1) Australian *C. edentula*, 2) Australian *C. maritima*; and 3) Australian hybrids. For western North America we had three groups: 1) western North American *C. edentula*; 2) western

North American *C. maritima*; and 3) western North American hybrids. No SNP blocking was used for TreeMix as the data set had been trimmed for linkage disequilibrium.

2.7 Genetic diversity and differentiation

Genetic diversity and differentiation within the two native ranges and two introduced ranges were assessed for the 256 individuals (the New Zealand and *C. lanceolata* samples were excluded) using the *global thinned dataset*. We calculated observed heterozygosity (H_O) and allelic richness (A_R). The 95% confidence intervals of A_R were calculated with 1000 bootstraps. These analyses were conducted using the diveRsity package (Keenan, McGinnity, Cross, Crozier, & Prodöhl, 2013). Because sampling at individual locations was limited in the native ranges, we grouped individuals based on their range, and their hybrid ancestry (pure parental or hybrid) using the *supervised run* Q-value assignments of the *global thinned dataset* into eight groups: 1) *C. edentula* from eastern North America; 2) *C. maritima* from Europe and northern Africa; 3) Australian *C. maritima*; 4) Australian *C. edentula*; 5) Australian hybrids; 6) western North American *C. maritima*; 7) western North American *C. edentula*; and 8) western North American hybrids. We used the Q value assignment of the *C. edentula* cluster and the highest standard error (0.024) of the *supervised run* to classify individuals. Individuals were considered hybrids if an individual had a $0.025 < Q > 0.975$ of the *C. edentula* cluster. To determine regional differentiation we calculated Weir and Cockerham's (1984) pairwise F_{ST} between the above eight groups using the *global thinned dataset* with VCFtools (Danecek et al., 2011). Additionally, we calculated the F_{ST} for pure parental individuals, grouping individuals according to their Admixture cluster from the *unsupervised run* and range (see Supplementary Information for more detail).

3 Results

3.1 Genetic structuring and differentiation

The Admixture analysis of the *unsupervised run* showed genetic structuring of *C. maritima*, *C. edentula* and hybrids. We plotted population pie charts and an individual bar plot for K=8 (Figure 1 A, B), which was the optimal K value. Genetic structure was present in the native range of *C. edentula,* where single samples from Lake Michigan and Rhode Island constituted one group, samples from New Brunswick within the Gulf of St. Lawrence the second group, samples from Newfoundland and Quebec (along the St Lawrence River) the third group and samples from Nova Scotia the final group. As expected, for *C. maritima,* there were two main groups: one group was largely from the Baltic and Atlantic coasts, which we term the "Atlantic" group (comprised mainly of the dark blue cluster, Figure 1 A, B) and a second admixed group was associated with the Mediterranean, that we term the "Mediterranean" group (comprised mainly of the light and medium blue clusters, Figure 1 A, B). In Australia, several genetic clusters were identified. First, in Queensland, New South Wales and Tasmania we identified pure *C. edentula* individuals with no evidence of hybridization with *C. maritima.* Second, for populations along the west coast of Australia, we identified a *C. maritima* cluster associated with the Atlantic coast in the native range. Third, in South Australia, genetic clusters associated with the Mediterranean were found. Finally, in the south-east of Australia there was evidence of hybrids between *C. maritima* and *C. edentula* (see below for details). In the introduced range of western North America, we identified pure *C. edentula* along with pure *C. maritima* (Figure 1 A, B). A small number of samples from Washington, Oregon and California showed evidence of hybridization (see hybrid classification section below).

The PCA and SPLITSTREE5 analyses confirmed the findings of Admixture. There was clear differentiation of *C. maritima* and *C. edentula* in the *global thinned dataset*. The first eigenvector

(EV) (Figure 2 A & S4 A) explained 33.17% of the variation and clearly delineated the species. The *C. edentula* group showed less variation than the *C. maritima* group along the first two EVs. Two *C. maritima* groupings were also evident with one representing *C. maritima* from Europe and Australia (EV1<0, EV2<0) and the other representing exclusively *C. maritima* from western North America (EV1<0, EV2>0). In the SPLITSTREE5 network, using the *global Splitstree dataset*, *C. edentula* (as identified by the *supervised run*) formed a monophyletic group without admixture. *C. maritima* samples were split into three groups (Figure 2 B, C): *C. maritima* associated with the Mediterranean group, *C. maritima* associated with an Atlantic group and *C. maritima* in western North America. Hybrids of the two species were scattered in between the *C. maritima* groups or between the two-parental species along the network. We conducted an additional native range SPLITSTREE5 analysis (Figure S5) that mirrors this pattern but provides clearer *C. edentula* grouping in the native range.

Pairwise F_{ST} (Table S2) using the *global thinned dataset* revealed clear genetic differentiation between the two-parental species originating from the native range ($F_{ST} > 0.527$). Within the introduced ranges the pairwise F_{ST} between the two species was similar to the comparison of the native ranges. Hybrids identified using Admixture in the introduced ranges showed higher genetic differentiation from *C. edentula* than from *C. maritima* (Table S2).

3.2 Genetic diversity

Population statistics revealed that in their native ranges *C. edentula*, the self-compatible species, has considerably less observed heterozygosity than *C. maritima* and the hybrids of the two species (Table S3). Allelic richness was significantly reduced in *C. edentula* in comparison to *C. maritima,* the largely self-incompatible species. In the introduced ranges, no clear reduction of H_O or A_R was observed in either of the species; indeed, *C. maritima* individuals seemed to have an increase in H_O

362 (in comparison to the native range). Hybrids of the two-species had higher H_O and A_R compared to
363 both parental species.

364

365 **3.3 Hybrid classification**

366 The three approaches classified different proportions of individuals as hybrids, as expected due to
367 their ability to detect recent hybrids (NewHybrid, HIest), versus hybrid ancestry (Admixture, HIest).
368 All hybrids identified by NewHybrids were also identified as hybrids with HIest and Admixture
369 (Table S4 & S5). The fourteen putative hybrids included in the samples as a result of morphological
370 identification were assigned by all analyses as hybrids, providing evidence of the accuracy of the
371 assignments. Furthermore, the NewHybrid analysis confirmed that these hybrids were likely the
372 product of the first two generations of interbreeding. NewHybrids analysis revealed 19 hybrids
373 (Figure 4; Table S4) with 17 hybrids in Australia (13.49%), one in western North America (1.47%)
374 and one in New Zealand. In Australia, F1 and F2 hybrids were detected in the current sympatric
375 zones where individuals with both species' phenotypic traits were clearly identifiable in the
376 populations. Hybrids (Figure S4 B) grouped in the PCA according to their generation, with F1 and F2
377 hybrids grouped between the parental species, and backcrosses grouped closer to species they
378 backcrossed to. In this same PCA the advanced generation hybrids identified with the *supervised run*
379 of Admixture as well as HIest frequently grouped with *C. maritima,* suggestive of further
380 backcrossing to that species.

381 Classification of hybrids using the *supervised run* of Admixture revealed 73 hybrids in Australia
382 (57.94%) from 15 locations, 11 hybrids in western North America (16.18%) from five locations and
383 one hybrid from New Zealand (Figure 1; Table S4). In western North America hybrids were found in
384 each of two locations in California and Oregon and in one location in Washington.

385 All Admixture hybrids were also identified as hybrids in HIest. When the 471 loci that are fixed
386 between native range samples were used, and we allowed for a low level of polymorphism within
387 each species (0.99 *C. edentula*, 0.03 *C. maritima*), a larger number of hybrids were identified using
388 HIest than Admixture (138 versus 85, Table S6). The additional hybrids identified by HIest over
389 Admixture were exclusively found in the introduced ranges and were identified as advance
390 generation hybrids with most showing a greater proportion of ancestry to *C. maritima* than *C.*
391 *edentula* (Figure 3). When we increased the allele frequency of *C. maritima* (0.99 *C. edentula*, 0.06
392 *C. maritima*) we identified slightly fewer hybrids (132). Again, they were all in the introduced
393 ranges. The discrepancy between runs mainly reflected the classification of individuals with an
394 apparent low level of ancestry from the alternate species. When 63 SNPs that were fixed between all
395 parental individuals based on the *supervised* Admixture analysis were used, the identification of
396 parental and hybrids was identical between methods Admixture and HIest methods. Furthermore, a
397 single F2 was identified (matching NewHybrid). In all the runs, advanced generation hybrids were
398 identified in this analysis with many in regions where *C. maritima* has not been recorded for many
399 decades, but also in the current sympatric zone (New South Wales, Queensland and Tasmania).

400 We then examined if patterns of ancestry in Australia and western North America reflected the likely
401 invasion route of *C. maritima*. Specifically, we tested if low levels of *C. edentula* ancestry were
402 found in areas where *C. maritima* first arrived, and if high levels of *C. edentula* ancestry were found
403 in regions *C. maritima* has more recently invaded and where *C. edentula* is still present. Using the
404 supervised Admixture analysis, the mean *C. edentula* ancestry of hybrids at each location was
405 correlated with the ranked distance from where *C. maritima* first arrived in south-eastern mainland
406 Australia ($\rho = 0.82$, $p < 0.01$) (Table 3). This pattern was also significant when testing across all
407 samples, including individuals identified as parental species ($\rho = 0.89$, $p < 0.05$). However, in
408 western North American, although the direction of the correlation was as predicted, a geographic

409 pattern in ancestry was only significant when using locations north of San Francisco as well as
410 parental and hybrid individuals ($\rho = 0.72$, $p < 0.05$). The same pattern of significance was found
411 when using the results of the HIest (Figure 3 & 5 & S6; Table 3)

412 To further confirm our finding of hybridization in the introduced ranges between the species, we used
413 TreeMix to assess geneflow between *C. edentula* and *C. maritima* within each introduced range. The
414 maximum likelihood tree in both invasive ranges showed bidirectional gene flow (Figure 6). In
415 Australia geneflow occurred from the *C. edentula* branch (comprised of eastern North American *C.*
416 *edentula* and Australian *C. edentula*) into Australian *C. maritima* (Mediterranean); a migration event
417 also occurred from this group into the *C. edentula* branch (Figure 6 B). In western North America the
418 same pattern occurs. There is evidence of a migration event from the *C. edentula* branch into western
419 North American *C. maritima* as well as a migration event from the western North American *C.*
420 *maritima* branch into the western North American *C. edentula* (Figure 6 A). We used the f_3 statistic of
421 TreeMix (Table 3) to further confirm hybridization within the introduced ranges. Hybrids (identified
422 by the *supervised* Admixture *run*) are admixed from the *C. edentula* and *C. maritima* parental
423 individuals within both introduced ranges (Australia $f_3 = -0.006$, $Z = -31.97$; western North America
424 $f_3 = -0.005$, $Z = -23.22$).

425

426 **4 Discussion**

427 Our analysis sheds light on the origin and extent of hybridization of two introduced species in two
428 separate invasions, which experienced a parallel pattern of invasion and apparent replacement of one
429 species by another. The *unsupervised run* of Admixture provides evidence that *C. edentula*
430 populations in Australia were likely from a single source, while in western North America *C.*
431 *edentula* likely originated from two different regions of eastern North America. *Cakile maritima* in

Australia was likely sourced from two distinct regions, with the western Australian populations originating from the European Baltic or Atlantic coasts and the south-eastern Australian populations from the Mediterranean. The western North American *C. maritima* populations likely originate from a single source and show the closest affinity to the Mediterranean samples in the Admixture analysis. Importantly, we found frequent hybridization in Australia (hybrid samples = 58%, *supervised run* Admixture) as well as the first genetic evidence of hybrids in western North America (16%, *supervised run* Admixture) and in New Zealand. In addition, the geographic distribution of hybrid ancestry fits with expectations based on historical records documenting the range expansion and replacement of *C. edentula* by *C. maritima*. Except at places where the two species are currently sympatric and new hybrids are still being formed, it would be difficult to determine morphologically that hybridization has ever taken place, since backcrossing soon hides its phenotypic evidence. *Cakile maritima* is highly variable within and between populations in its native range and hybrids in the introduced range could easily be overlooked (e.g. Cousens et al., 2013) without the use of molecular methods. It is therefore an intriguing possibility that hybridization may be commonly overlooked in a much wider range of invasive taxa, especially where morphological trait indicators of hybridization are more cryptic. Alien floras commonly include many congeneric species whose capacity for interbreeding is yet to be established. While previous authors (Ellstrand & Schierenbeck, 2000) have raised our attention to obvious hybrid species and allopolyploids, perhaps the impacts of hybridization are often more insidious. It is thus important – though not an easy task – to determine in future the extent to which such non-apparent introgression has been beneficial during invasion.

4.1 Native range patterns

One of our primary goals was to identify the source regions for the invasions for each species, which is only possible when there is geographic structuring in the native ranges. Our analysis provided evidence of geographic structuring in the *C. edentula* native range, at a much finer grain than

456 currently recognized taxonomically (Figure 1). Samples from Quebec, Newfoundland, Nova Scotia
457 and New Brunswick contain separate Admixture clusters, likely within *C. edentula* subsp. *edentula*
458 var. *edentula* as this subspecies is the only one described in this region of the North American
459 Atlantic coast (Rodman, 1974). Two single samples from Lake Michigan and Rhode Island grouped
460 together in one cluster of the Admixture analysis; those samples might belong to the Atlantic coast
461 variety of *C. edentula* subsp. *edentula* var. *edentula* as it is known to have invaded Lake Michigan in
462 historical times (Huebner, 2009; Rodman, 1974), where it now coexists with the Great Lakes
463 endemic var. *lacustris*. A second possibility, suggested by Gormally et al., (2011), but without
464 morphological evidence, is that var. *lacustris* has dispersed to the Atlantic. Genetically distinct
465 regional variation is not surprising, as the directions of currents and the influences of geological
466 features on seed dispersal can be highly predictable (Lapointe, 2000). Similar conclusions have been
467 reached in the Mediterranean by Westberg (2005) and Gandour et al. (2008). *Cakile edentula* subsp.
468 *harperi* occurs in areas south of the populations sampled in our study (Rodman, 1974), but
469 comprehensive studies of herbarium samples by Rodman (1974) and Cousens et al., (2013) have
470 found no morphological evidence that subsp. *harperi* has been introduced anywhere outside its native
471 range.

472 Our analyses revealed clustering of *C. maritima* in its native Europe largely consistent with the
473 accepted taxonomic distributions (Ball, 1964; Marhold, 2011; Rodman, 1974) as well as one previous
474 population genetic analysis (Clausing et al., 2000). Other genetic studies with greater sampling
475 intensity, however, showed more differentiation on a local level (Kadereit et al., 2005; Westberg,
476 2005). The absence of fine-grain local differentiation in our study might be driven by the limited
477 number of native range samples for this species and restricted sampling of the Baltic area.

478 *Cakile edentula* showed lower genetic diversity than *C. maritima* in their native ranges as measured
479 by allelic richness and observed heterozygosity (Table S3) and showed less variation along the EVs

and in the SPLITSTREE network analysis (Figure 2). Higher selfing rates in *C. edentula* would be expected to reduce the effective population size compared to the largely self-incompatible *C. maritima* (Pollak, 1987).

4.2 Introduced range patterns

4.2.1 Australia and New Zealand

Although *C. edentula* has now disappeared from much of its original introduced range in Australia, some pure *C. edentula* populations still remain. Our analyses show that they likely originate from populations located in Nova Scotia as they contained an Admixture cluster found exclusively in this region of the native range and showed the lowest genetic differentiation from this region (Figure 1; Table S7). *Cakile edentula* A_R and H_O did not change considerably in Australia compared to the native range (Table S3), which is inconsistent with a strong invasion bottleneck. The genetic structure of the Australian *C. maritima* samples is consistent with a history of multiple introductions. This is in accordance with previous morphological and genetic studies of invasion history in Australia (Cousens et al., 2013; Ohadi et al., 2016; Rodman, 1976, 1986). In particular, the cluster associated with the Atlantic European group is found in western Australia, while a Mediterranean cluster predominates in southern and eastern Australia (including Tasmania) (Figure 1; Table S8). Similarly, analysis of microsatellite markers indicated that that western and south-eastern populations of *C. maritima* in Australia were genetically distinct and most likely resulted from independent introductions with severely limited gene flow from west to east (Ohadi et al., 2016). Finally, Australian *C. maritima* showed higher A_R and H_O values than its native range, consistent with admixture of multiple source populations and/or hybridization with *C. edentula.* Many successful invasions are sourced from multiple introductions (e.g., Vallejo- Marín et al., 2020; van Boheemen et al. 2017) and both hybridization and multiple introductions and admixture may spur successful

503 invasions (Ellstrand & Schierenbeck, 2006; Dlugosch & Parker, 2008; Hodgins, Bock, Rieseberg,
504 2018).

505 Our data provides substantial evidence for extensive hybridization in Australia between the two
506 species. TreeMix supported bidirectional gene flow between Australian *C. maritima* and *C. edentula*
507 individuals (identified morphologically) (Figure 6). This was confirmed by the Admixture global
508 analysis (Figure 1), the PCA and Splitstree analysis, as many Australian samples fell in-between the
509 native range samples of both species (Figure 2), and the f_3 test (Table 3). Further support is provided
510 by three separate analyses which specifically detect hybrid individuals (Figure 1 & 3 & 5 & S6;
511 Table S4). As expected, Australian hybrids (*supervised* Admixture *run*) had higher genetic diversity
512 than both parental species (Table S3). Furthermore, the pattern of hybrid ancestry was geographically
513 structured and reflected the historical invasion route of *C. maritima* in south-eastern Australia. This
514 pattern was consistent across two separate approaches (*supervised* Admixture *run*, HIest) to identify
515 hybrid ancestry (Figure 1 & 3; Table 2). NewHybrids confirmed the presence of a small number of
516 early generation hybrids (within two generations) where both species still co-occur (Figure 4). Some
517 mixed populations in Australia show pure genotypes of both parental species and early generation
518 hybrids, demonstrating on-going hybridization of the two taxa (Figure 4). In areas where *C. edentula*
519 still persists, backcrossing to *C. edentula* has also occurred, but is rare, and recent backcrosses to *C.*
520 *maritima* appear to be more common. In those parts of Australia where *C. maritima* has already
521 appeared to have replaced *C. edentula* (i.e., where no *C. edentula* phenotypes remain; Cousens et al.,
522 2013; Rodman, 1986), evidence is consistent with past hybridization between the species and
523 repeated backcrossing to *C. maritima* (Figure 1 & 3 & 6). In areas of Western Australia, where *C.*
524 *edentula* has never been identified, evidence of hybridization with *C. edentula* was also found,
525 confirming a previous observation by Ohadi et al., (2016). The sample from New Zealand was
526 identified as a hybrid where the same replacement of *C. edentula* by *C. maritima* has also taken place

527 (Cousens & Cousens, 2011).

528 **4.2.2 Western North America**

529 Our results revealed that *C. edentula* in western North America most likely originated from two
530 sources in eastern North America. We also found that western North American *C. maritima*
531 potentially originated from the Mediterranean region, as *C. maritima* in western North America
532 contained the same Admixture clusters as the Mediterranean and showed the lowest differentiation
533 from this region (Figure 1; Table S7 & S8). However, these populations were genetically distinct
534 (Figure 2 & S4) suggesting the possibility of an unknown source for this invasion, or the impact of an
535 invasion bottleneck. *Cakile edentula* and *C. maritima* in western North America showed, as in
536 Australia, no reduction H_O and A_R, which may reflect the impacts of undetected hybridization, large
537 founding populations, or multiple introductions.

538 Like Australia, hybridization was identified between the two species in western North America,
539 although the proportion of hybrids was less (e.g., 58% versus 16% using the *supervised Admixture*
540 *run*). TreeMix identified bidirectional gene flow between the species in western North America
541 (Figure 6; Table 3), and evidence consistent with hybridization was apparent in the global Admixture
542 analysis (Figure 1), the PCA and Splitstree analysis (Figure 2). Furthermore, we employed three
543 independent methods to specifically identify hybrid individuals and their likely generation. From this
544 we identified 11 hybrid samples (all 11 were identified by both HIest and Admixture and one as an
545 F2 by NewHybrids) from five locations in western North America. Specimens of hybrids based on
546 morphological identification are largely unknown for this region, either in herbaria or in the field
547 (Rodman, 1974). But more recently, Cody and Cody (2004) reported a small percentage of hybrids in
548 a population from British Columbia. Although the fitness and demographic consequences of
549 hybridization during introduction require further investigation, the lower incidence of hybrids in

western North America compared to Australia suggests that hybridization could have facilitated the establishment and rapid spread of *C. maritima* to a greater degree in Australia. In support of this hypothesis, the complete replacement of *C. edentula* by *C. maritima* phenotypes has not progressed as far north in western North America compared to Australia, where few northern populations of *C. edentula* remain. However, the mechanism driving differences in hybridization rates in western North America compared to Australia is unclear and requires further investigation.

4.3 Hybrid identification and significance

The pattern of invasion first by *C. edentula*, then by *C. maritima*, has been repeated in three regions. Prior to this study, hybrids were known only from Australia. However, we also identified clear evidence of hybridization in western North America and in New Zealand. Hybrids between the two species can be produced readily by handcrossing (e.g. Li et al., 2019; Mesgaran et al., 2016; Rodman, 1974) and our data demonstrate that recent and advanced generation hybrids are at least partially fertile in natural populations. Our results show backcrossing to both parental species, although backcrossing to *C. maritima* was much more frequent (Figure 3). This pattern of biased backcrossing towards *C. maritima* was predicted based on field observations of pollinator visitations (Mesgaran et al., 2016), the morphological replacement of *C. edentula* by *C. maritima,* and previous genetic studies (Mesgaran et al., 2016; Ohadi et al., 2016). It is also consistent with expected mating asymmetries between these species and their hybrids caused by the inheritance of the self-incompatibility system and traits associated with pollinator attraction in hybrids (Li et al., 2019). In artificial crosses, early generation hybrids inherited mostly (but not exclusively) self-incompatibility, as well as larger floral displays, similar to *C. maritima* (Li et al., 2019). This suggests that F1 hybrids will often need to rely on outcrossing, and that larger floral displays should facilitate this. Consequently, these traits in the hybrids should further contribute to backcrossing to the self-incompatible parent (*C. maritima*). A similar asymmetric pattern of species ancestry has been

identified other hybrids of other species with such differences in mating system (Brandvain, Kenney, Flagel, Coop, & Sweigart, 2014; Pickup et al., 2019; Ruhsam, Hollingsworth, & Ennos, 2011).

Our identification of advanced generation backcrosses to *C. maritima* means that portions of the *C. edentula* genome have been retained in a largely *C. maritima* background (i.e. introgression), long after morphological evidence of hybridization has gone from a population. The role of selection and neutral evolutionary processes in governing patterns of introgression across the genome, however, remains to be investigated in this system. Theory suggests that regions of the genome that are not introgressed will harbour incompatibilities or a high number of additive deleterious alleles in the introgressing species (Harris & Nielsen, 2016; Juric, Aeschbacher, & Coop 2016). A greater fixation rate of weakly deleterious alleles is predicted in the *C. edentula* due to its higher level of inbreeding, and indeed, the low levels of genetic variability in this species relative to *C. maritima* support a lower effective population size in this species. Selection against a higher genetic load originating from *C. edentula* in hybrids should more rapidly lead to the reconstitution of a *C. maritima* genome following transient hybridization during range expansion. In line with the expectation of selection against selfing ancestry in outcrossers, in *Mimulus gutattus* (outcrossing) genomic regions with high recombination rates have reduced levels of ancestry from the selfing species *Mimulus nasutus* (Brandvain et al., 2014). However, several remarkable examples in plants have demonstrated the infusion of favorable alleles via hybridization (adaptive introgression), including the transfer of herbivore resistance in *Helianthus* (Whitney, Randell, & Rieseberg, 2006). Indeed, Cody and Cody (2004) proposed the intriguing possibility of adaptive introgression in this system but this remains to be investigated. Our identification of replicated patterns of hybridization, replacement and invasion in *Cakile* provide an exciting opportunity for further investigation of the beneficial and detrimental consequences of hybridization during range expansion.

5 Conclusion

For more than 40 years the mechanism by which an established invader (*C. edentula*) has been replaced by a subsequent introduced species (*C. maritima*) in three separate parts of the world has remained a puzzle (Barbour & Rodman, 1970). Here we confirm that, particularly in Australia, the apparent replacement of *C. edentula* by *C. maritima* is not complete and remnants of the *C. edentula* genome are evident in contemporary *C. maritima* populations. Furthermore, it appears that both early and later generation hybrids are at least partially fertile in natural populations and that there is a higher frequency of backcrossing to *C. maritima*. The patterns of hybridization we identified is consistent with the hypothesis that mating among these cross-compatible invaders has facilitated the establishment of the self-incompatible *C. maritima* whose range expansion may otherwise be limited due to Allee effects, as has been observed in other potential self-incompatible invaders (Uesugi, Baker, de Silva, Nurkowski, & Hodgins, 2020). The demographic benefits to *C. maritima* of hybridization during range expansion have been assessed through simulations (Mesgaran, et al. 2016). However further experimental studies examining Allee effects in this self-incompatible species, and whether mixed-species populations can mitigate these effects, are needed. Likewise, the evolutionary consequence of hybridization for both species remains unclear, as is its role, if any, in the rapid expansion of one invader at the expense of another.

6 Author Contributions

KH, RC and LR conceived of and designed the study. KH, KN and RC carried out sampling. KN conducted the molecular laboratory work. HR carried out the bioinformatics analyses with significant input from AG, PB and KH. AG, KH, LR, PB, RC and HR contributed to the writing and approved the final manuscript.

7 Funding

KH, RC and LR received funding from the Australian Research Council (Grant ID: DP180102531). HR was supported by Monash Graduate Scholarship (MGS).

8 Acknowledgments

We would like to thank all collectors of this study. We also would like to thank Sarah Bou-assi for help with molecular laboratory work and Lotte van Boheemen for help with the initial bioinformatics analysis.

9 Data Availability Statement

The datasets generated and analyzed for this study can be found in the Sequence Read Archive (SRA) of GenBank. [http://www.ncbi.nlm.gov/bioproject/637114]. Scripts available on https://github.com/HannaRos/Cakile-GBS-scripts.

10 Data reference

[dataset]. Rosinger, H.S., Geraldes, A. M., Nurkowski, K. A., Battlay, P., Cousens, R. D., Rieseberg, L. H., Hodgins, K. A.; 2020; GBSCAK.vcf.gz; Monash University Bridges; DOI: 10.26180/5ef01e7a6359b

[dataset]. Rosinger, H.S., Geraldes, A. M., Nurkowski, K. A., Battlay, P., Cousens, R. D., Rieseberg, L. H., Hodgins, K. A.; 2020; filtered dataset GBS Cakile; Monash University Bridges; DOI:10.26180/5f6be2a84dc4d

11 References

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

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655-1664. https://doi.org/10.1101/gr.094052.109

Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, *160*(3), 1217-1229.

Baack, E., Melo, M. C., Rieseberg, L. H., & Ortiz−Barrientos, D. (2015). The origins of reproductive isolation in plants. *New Phytologist*, *207*(4), 968-984. https://doi.org/10.1111/nph.13424

Baker, H. G. (1965) Characteristics and modes of origin of weeds. In: The Genetics of Colonizing Species (eds Baker H, Stebbins G), pp. 147–168. Academic Press, New York.

Baker, H. G. (1974). The evolution of weeds. *Annual Review of Ecology and Systematics*, *5*(1), 1-24. https://doi.org/10.1146/annurev.es.05.110174.000245

Ball, P. W. (1964). A revision of *Cakile* in Europe. *Feddes Repertorium*, *69*, 35-40.

Barbour, M. G. (1970). Seedling ecology of *Cakile maritima* along the California coast. *Bulletin of the Torrey Botanical Club*, 280-289. https://doi.org/10.2307/2483647

Barbour, M. G., & Rodman, J. E. (1970). Saga of the West Coast sea-rockets: *Cakile edentula* ssp. *californica* and *C. maritima*. *Rhodora*, *72*(791), 370-386.

Bock, C.H. (2008) The effect of *Alternaria brassicicola* infection on the reproductive fitness of the naturally occurring littoral ruderals *Cakile maritima* and *C. edentula. Australian Plant Pathology* *37*(6), 569-580. https://doi.org/10.1071/AP08057

Bock, D. G., Caseys, C., Cousens, R. D., Hahn, M. A., Heredia, S. M., Hübner, S., ... Rieseberg, L. H. (2015). What we still don't know about invasion genetics. *Molecular Ecology*, *24*(9), 2277-2297. https://doi.org/10.1111/mec.13032

Brandvain, Y., Kenney, A. M., Flagel, L., Coop, G., & Sweigart, A. L. (2014). Speciation and Introgression between *Mimulus nasutus* and *Mimulus guttatus*. PLoS Genet 10(6) : e1004410. https://doi:10.1371/journal.pgen.1004410

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, *1*(3), 171-182. https://doi.org/10.1534/g3.111.000240

Boyd, R. S., & Barbour, M. G. (1993). Replacement of *Cakile edentula* by *C. maritima* in the strand habitat of California. *American Midland Naturalist*, 209-228. https://doi.org/10.2307/2426122

Clausing, G., Vickers, K., & Kadereit, J. W. (2000). Historical biogeography in a linear system: genetic variation of Sea Rocket (*Cakile maritima*) and Sea Holly (*Eryngium maritimum*) along European coasts. *Molecular Ecology*, *9*(11), 1823-1833. https://doi.org/10.1046/j.1365-294x.2000.01083.x

Cody, T. W., & Cody, M. L. (2004). Morphology and spatial distribution of alien sea-rockets (*Cakile spp.)* on South Australian and Western Canadian beaches. *Australian Journal of Botany*, *52*(2), 175-183.

Cousens, R. D., Ades, P. K., Mesgaran, M. B., & Ohadi, S. (2013). Reassessment of the invasion history of two species of *Cakile* (Brassicaceae) in Australia. *Cunninghamia*, *13*, 275-290.
Cousens, R. D., & Cousens, J. M. (2011). Invasion of the New Zealand coastline by European sea-rocket (*Cakile maritima*) and American sea-rocket (*Cakile edentula*). *Invasive Plant Science and Management*, *4*(2), 260-263.
Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158. https://doi.org/10.1093/bioinformatics/btr330
Dlugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology, 17*(1), 431-449. doi:10.1111/j.1365-294X.2007.03538.x
Ellstrand, N. C., & 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. https://doi.org/10.1073/pnas.97.13.7043
Fitzpatrick, B. M. (2012). Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evolutionary Biology*, *12*(1), 131. https://doi.org/10.1186/1471-2148-12-131
Fox, J. & Weisberg S. (2019). An R companion to applied regression, third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/
Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27-32. https://doi.org/10.1111/1755-0998.12509
Gandour, M., Hessini, K., & Abdelly, C. (2008). Understanding the population genetic structure of coastal species (*Cakile maritima*): seed dispersal and the role of sea currents in determining population structure. *Genetics Research*, *90*(2), 167-178. https://doi.org/10.1017/S0016672308009269
Gormally, C. L., Hamrick, J.L., & Donovan, J. H. A. (2011). Genetic structure of a widely dispersed beach annual, *Cakile edentula* (Brassicaceae). *American Journal of Botany*, *98*(10), 1657-1662. https://doi.org/10.3732/ajb.1000499
Harris, K., & Nielsen, R. (2016). The genetic cost of Neanderthal introgression. *Genetics*, *203*(2), 881-891. https://doi.org/10.1534/genetics.116.186890
Harrison, B. J., & Darby, L. A. (1955). Unilateral hybridization. *Nature*, *176*(4490), 982-982. https://doi.org/10.1038/176982a0
Heyligers, P. C. (1984). Beach invaders: sea rockets and beach daisies thrive. *Australian Natural History*; 21(5),212-214
Hodgins, K. A., Bock, D. G., & Rieseberg, L. H. (2018). Trait evolution in invasive species. *Annual Plant Reviews Online*, 459-496. https://doi.org/10.1002/9781119312994.apr0643
Hoffmann, B. D., & Broadhurst, L. M. (2016). The economic cost of managing invasive species in Australia. *NeoBiota*, *31*, 1. https://doi.org/10.3897/neobiota.31.6960
Hovick, S. M., & Whitney, K. D. (2014). Hybridisation is associated with increased fecundity and size in invasive taxa: meta−analytic support for the hybridisation−invasion hypothesis. *Ecology Letters*, *17*(11), 1464-1477. https://doi.org/10.1111/ele.12355
Huebner, D. C. (2009). A morphological, molecular and experimental assessment of the conservation status of great lakes sea-rocket (*Cakile edentula var. lacustris,* Brassicaceae) (Unpublished doctoral dissertation). Northwestern University.
Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254-267. https://doi.org/10.1093/molbev/msj030
Joshi, N. A. and Fass, J. N. (2011). Sickle: a Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files (Version 1.33) [Software]. Retrieved from https://github.com/najoshi/sickle

Juric, I., Aeschbacher, S., & Coop, G. (2016). The strength of selection against Neanderthal introgression. *PLoS Genetics*, *12*(11), e1006340. https://doi.org/10.1371/journal.pgen.1006340
Kadereit, J. W., Arafeh, R., Somogyi, G., & Westberg, E. (2005). Terrestrial growth and marine dispersal? Comparative phylogeography of five coastal plant species at a European scale. *Taxon*, *54*(4), 861-876. https://doi.org/10.2307/25065567
Kassambara, A. (2020). Ggpubr: ggplot2 Based Publication Ready Plots. R package version 0.2.5 [R package]. A Retrieved from https://cran.r-project.org/package=ggpubr
Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). DiveRsity: An R package for the estimation of population genetics parameters and their associated errors. Methods Ecoly Evolution, *4*(8), 782-788. doi:10.1111/2041-210X.12067
Lapointe, M. (2000). Modern diatom assemblages in surface sediments from the Maritime Estuary and the Gulf of St. Lawrence, Québec (Canada). *Marine Micropaleontology*, *40*(1-2), 43-65. https://doi.org/0.1016/S0377-8398(00)00031-1
Li, C., Cousens, R. D., & Mesgaran, M. B. (2019). How can natural hybridisation between self−compatible and self−incompatible species be bidirectional?. *Weed Research*, *59*(5), 339-348. https://doi.org/10.1111/wre.12372
Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows- Wheeler transform. *Bioinformatics* 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324
Li, C., Mesgaran, M. B., Ades, P. K., & Cousens, R. D. (2020). Inheritance of breeding system in *Cakile* (Brassicaceae) following hybridization: implications for plant invasions. *Annals of Botany*, *125*(4), 639-650. https://doi.org/10.1093/aob/mcz198
Maddison, W. P. and Maddison, D.R. (2019). Mesquite: A Modular System for Evolutionary Analysis [Software]. Retrieved from http://www.mesquiteproject.org
Marhold, K. (2011). Brassicaceae. In: Euro+Med Plantbase- the Information Resource for Euro-Mediterranean Plant Diversity. Retrieved from http://ww2.bgbm.org/EuroPlusMed/PTaxonDetail.asp?NameId=14225&PTRefFk=7200000
Mesgaran, M. B., Lewis, M. A., Ades, P. K., Donohue, K., Ohadi, S., Li, C., & Cousens, R. D. (2016). Hybridization can facilitate species invasions, even without enhancing local adaptation. *Proceedings of the National Academy of Sciences*, *113*(36), 10210-10214. https://doi.org/10.1073/pnas.1605626113
Ohadi, S., Ades, P. K., Ford, R., Strand, A. E., Tibbits, J., Mesgaran, M. B., & Cousens, R. D. (2016). Genetic structure along the strandline: Unravelling invasion history in a one−dimensional system. *Journal of Biogeography*, *43*(3), 451-460. https://doi.org/10.1111/jbi.12640
Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337-2360. https://doi.org/10.1111/mec.13557
Pickrell, J.K., & Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, In Press.
Pickup, M., Brandvain, Y., Fraïsse, C., Yakimowski, S., Barton, N. H., Dixit, T., ... Field, D. L. (2019). Mating system variation in hybrid zones: facilitation, barriers and asymmetries to gene flow. *New Phytologist*, *224*(3), 1035-1047. https://doi.org/10.1111/nph.16180
Pimentel, D., Zuniga, R., & Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, *52*(3), 273-288. https://doi.org/10.1016/j.ecolecon.2004.10.002
Pollak, E. (1987). On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics*, *117*(2), 353-360.
Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., ... Shakir, K. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178. https://doi.org/10.1101/201178

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*(7263), 489-494. https://doi.org/10.1038/nature08365
Rodman, J. E. (1974). Systematics and evolution of the genus *Cakile* (Cruciferae). *Contributions from the Gray Herbarium of Harvard University*, (205), 3-146.
Rodman, J. E. (1976). Differentiation and migration of *Cakile* (Cruciferae): seed glucosinolate evidence. *Systematic Botany*, 137-148.
Rodman, J. E. (1986). Introduction, establishment and replacement of sea-rockets (*Cakile*, Cruciferae) in Australia. *Journal of Biogeography*, 159-171. https://doi.org/10.2307/2844990
RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA [Software]. Retrieved from http://www.rstudio.com/
Ruiz, G. M., Rawlings, T. K., Dobbs, F. C., Drake, L. A., Mullady, T., Huq, A., & Colwell, R. R. (2000). Global spread of microorganisms by ships. *Nature*, *408*(6808), 49-50. https://doi.org/10.1038/35040695
Ruhsam, M., Hollingsworth, P. M., & Ennos, R. A. (2011). Early evolution in a hybrid swarm between outcrossing and selfing lineages in *Geum*. *Heredity*, *107*(3), 246-255. https://doi.org/10.1038/hdy.2011.9
Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A., ... McCauley, D. E. (2001). The population biology of invasive species. *Annual review of Ecology and Systematics*, *32*(1), 305-332. https://doi.org/10.1146/annurev.ecolsys.32.081501.114037
Sax, D. F., & Gaines, S. D. (2003). Species diversity: from global decreases to local increases. *Trends in Ecology & Evolution*, *18*(11), 561-566.
Simberloff, D. (2013). Biological invasions: Prospects for slowing a major global change. *Elementa Science of the Anthropcene*, *1*, p.000008. doi: http://doi.org/10.12952/journal.elementa.000008
Sedlazeck, F. J., Rescheneder, P., & von Haeseler, A. (2013). NextGenMap: fast and accurate read mapping in highly polymorphic genomes. Bioinformatics, 29(21), 2790-2791. https://doi.org/10.1093/bioinformatics/btt468
Thrall, P. H., Young, A. G., & Burdon, J. J. (2000). An analysis of mating structure in populations of the annual sea rocket, *Cakile maritima* (Brassicaceae). *Australian Journal of Botany*, *48*(6), 731-738. https://doi.org/10.1071/BT99060
Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., ... Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, *9*(7), 892-908. https://doi.org/10.1111/eva.12367
Uesugi, A., Baker, D. J., de Silva, N., Nurkowski, K., & Hodgins, K. A. (2020). A lack of genetically compatible mates constrains the spread of an invasive weed. *New Phytologist*, *226*(6), 1864-1872. https://doi.org/10.1111/nph.16496
Vallejo-Marín, M., Friedman, J., Twyford, A. D., Lepais, O., Ickert-Bond, S. M., Streisfeld, M. A., ... & Puzey, J. R. (2020). Population genomic and historical analysis reveals a global invasion by bridgehead processes in *Mimulus guttatus*. *bioRxiv*. doi: https://doi.org/10.1101/2020.06.26.173286
van Boheemen, L. A., Lombaert, E., Nurkowski, K. A., Gauffre, B., Rieseberg, L. H., & Hodgins, K. A. (2017). Multiple introductions, admixture and bridgehead invasion characterize the introduction history of *Ambrosia artemisiifolia* in Europe and Australia. *Molecular Ecology*, *26*(20), 5421-5434. https://doi.org/10.1111/mec.14293
Vilatersana, R., Sanz, M., Galian, A., & Castells, E. (2016). The invasion of *Senecio pterophorus* across continents: multiple, independent introductions, admixture and hybridization. *Biological invasions*, *18*(7), 2045-2065. https://doi.org/10.1007/s10530-016-1150-1
Ward, S. M., Gaskin, J. F., & Wilson, L. M. (2008). Ecological genetics of plant invasion: what do we know?. *Invasive Plant Science and Management*, *1*(1), 98-109. https://doi.org/10.1614/IPSM-07-022.1

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. Evolution, 38, 1358–1370. https://doi.org/10.2307/2408641

Westberg, E. D. (2005). European phylogeography of the coastal plants *Cakile maritima Scop.* (Brassicaceae) and *Eryngium maritimum L.* (Apiaceae) (Unpublished doctoral dissertation). Johannes Gutenberg-Universität.

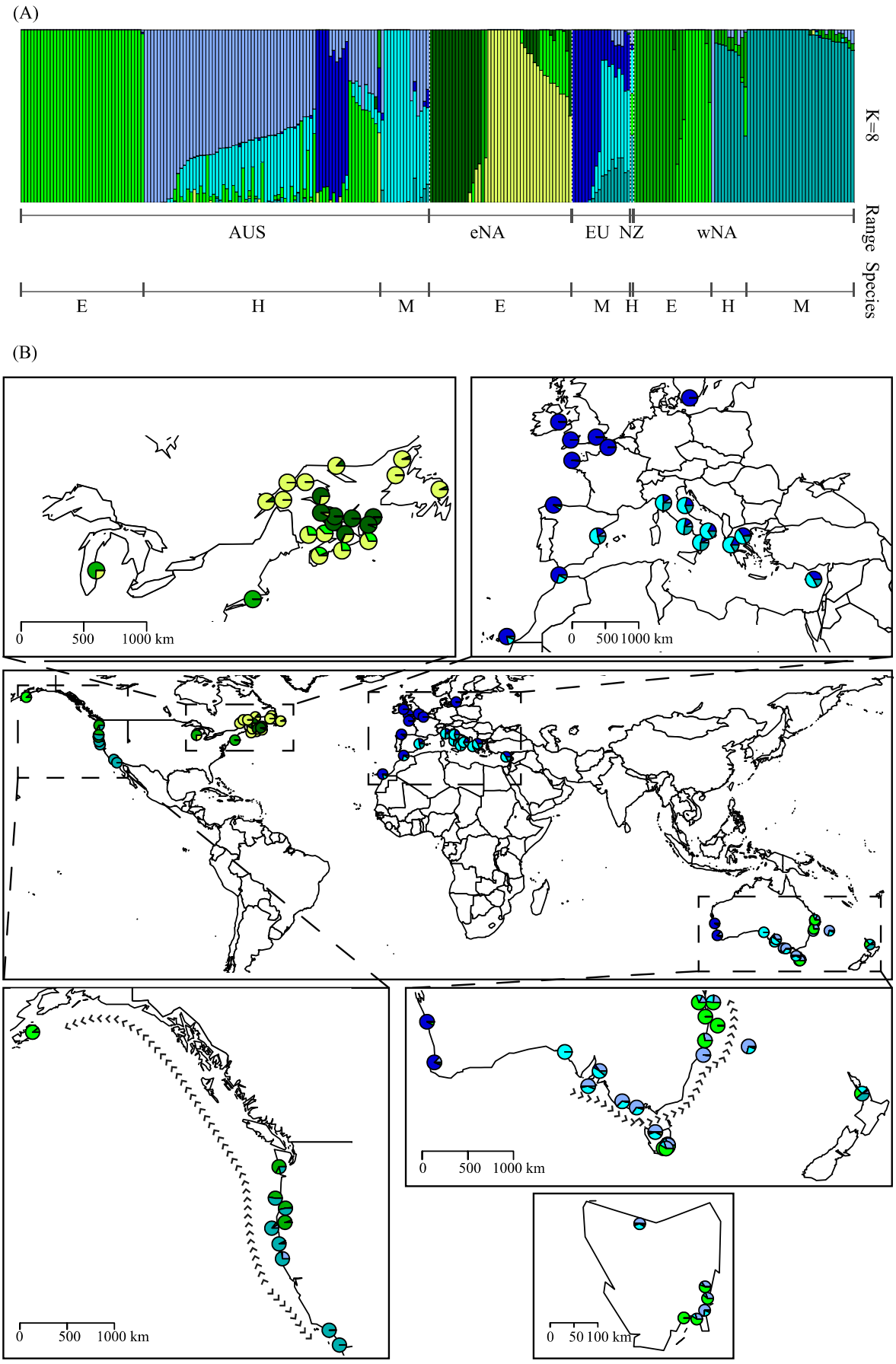
Whitlock, R., Hipperson, H., Mannarelli, M., & Burke, T. (2008). A high−throughput protocol for extracting high−purity genomic DNA from plants and animals. *Molecular Ecology Resources*, *8*(4), 736-741. https://doi.org/10.1111/j.1755-0998.2007.02074.x

Whitney, K. D., Randell, R. A., & Rieseberg, L. H. (2006). Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *The American Naturalist*, *167*(6), 794-807. https://doi.org/10.1086/504606

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., ... Kuhn, M. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326-3328. https://doi.org/10.1093/bioinformatics/bts606

Figures and figure captions


(A)
K=8
AUS
eNA
EU
NZ
wNA
Range
E
H
M
E
M
H
E
H
M
Species
(B)
0 500 1000 km
0 500 1000 km
0 500 1000 km
0 500 1000 km
0 50 100 km

838

Figure 1 Admixture results of the *unsupervised run* of the *global thinned dataset*. (A) A distruct plot for K=8. Individuals are ordered according to their cluster association of the *supervised run*. AUS=Australia, eNA= eastern North America, EU= Europe and northern Africa, NZ= New Zealand, wNA=western North America. E= *C. edentula*, M= *C. maritima*, H= Hybrids. (B) Population pie charts for K=8, Admixture proportions for each population are displayed. A global map is displayed as well as close ups of western North America, Europe, the Australian mainland and Tasmania. Colours correspond to the clusters in the distruct plot. Arrows indicates direction of invasion and direction of Spearman test.

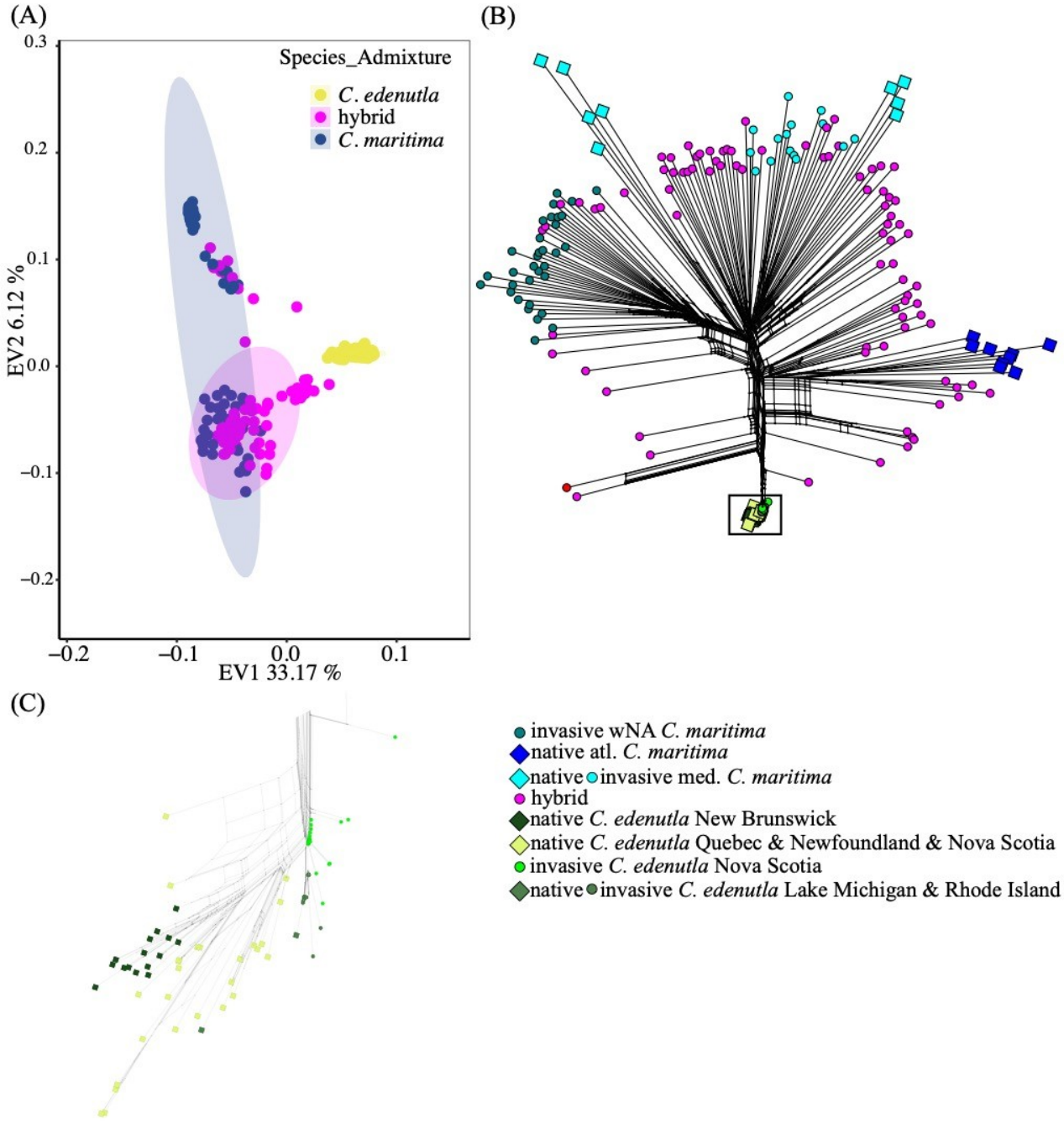
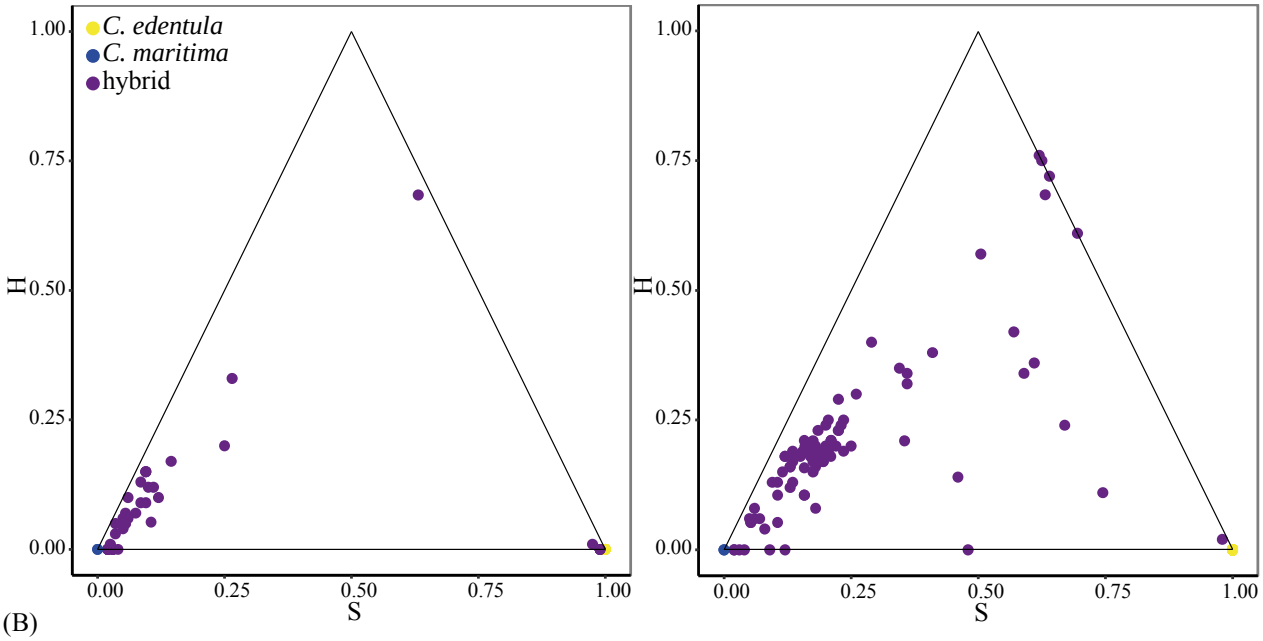


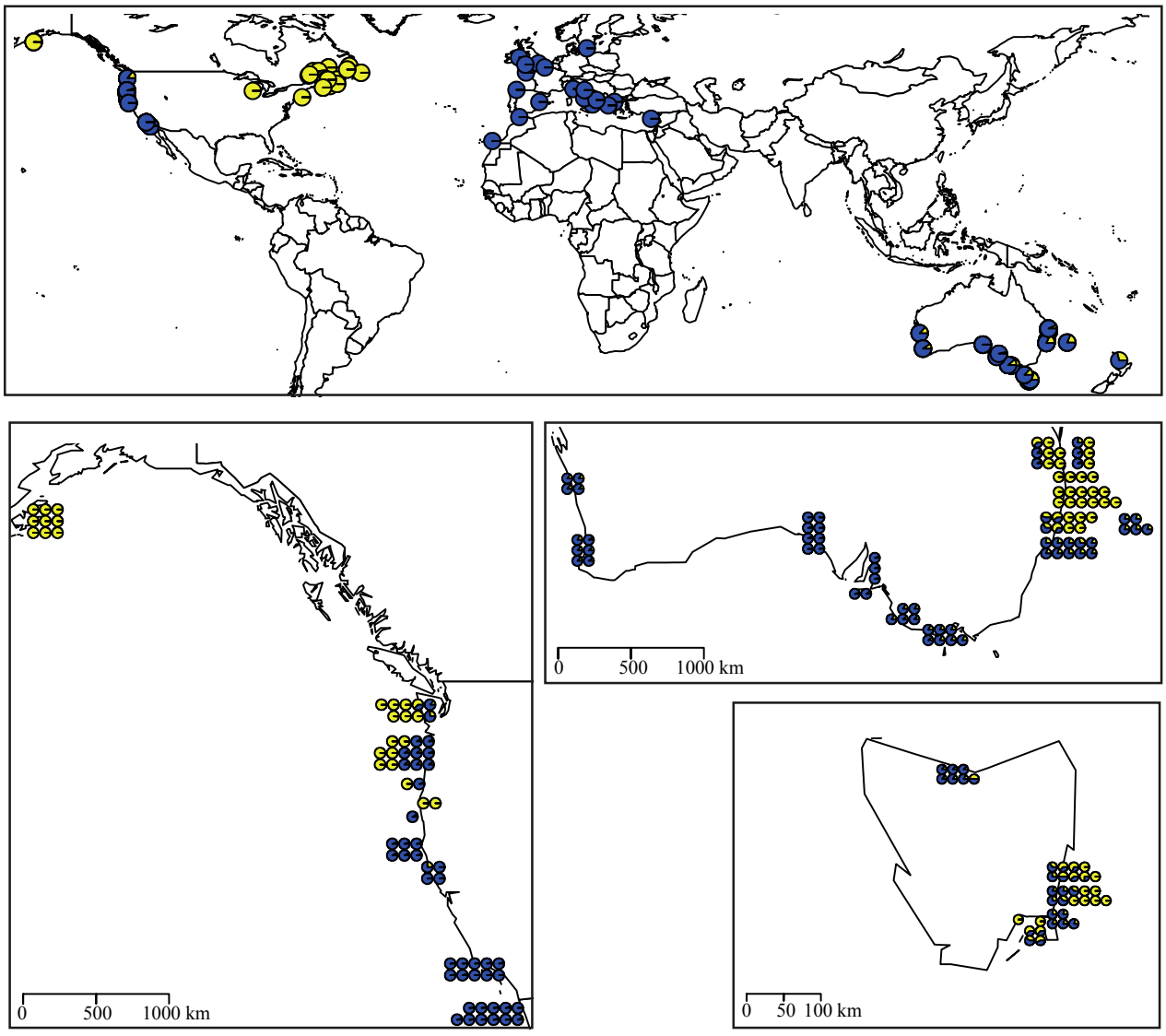
Figure 2 (A) Principal component analysis of the *global thinned dataset*. First two eigenvectors are presented. Individuals are coloured according to their species and hybrid status based on the

supervised run of Admixture. Ellipses indicate the 95% confidence range of the cluster. (B) Splitstree network of the *global Splitstree dataset*. Individuals are coloured according to their predominant cluster of the *unsupervised run* of Admixture cluster (K=8 of the *global thinned dataset*), with hybrids identified using the *supervised* run. The shapes indicate native vs. invasive range.

(A)


C. edentula
C. maritima
hybrid
H
S
1.00
0.75
0.50
0.25
0.00

(B)


0 500 1000 km
0 500 1000 km
0 50 100 km

854

Figure 3 Results of a hybridization assignment test implemented by HIest using 471 SNPs (0.99, 0.03). (A) Association of ancestry index (S) and interclass heterozygosity (H) are given for western North America (left) and Australia (right). Individuals are coloured according to their HIest classification. For hybrids the continuous model was a better fit than the hybrid classes. (B) The geographic distribution of individuals and their S index; yellow= *C. edentula* proportion, blue= *C. maritima* proportion. A global map and close-ups of western North America, the Australian mainland and Tasmania are presented. Arrows indicates direction of invasion and direction of Spearman's test.

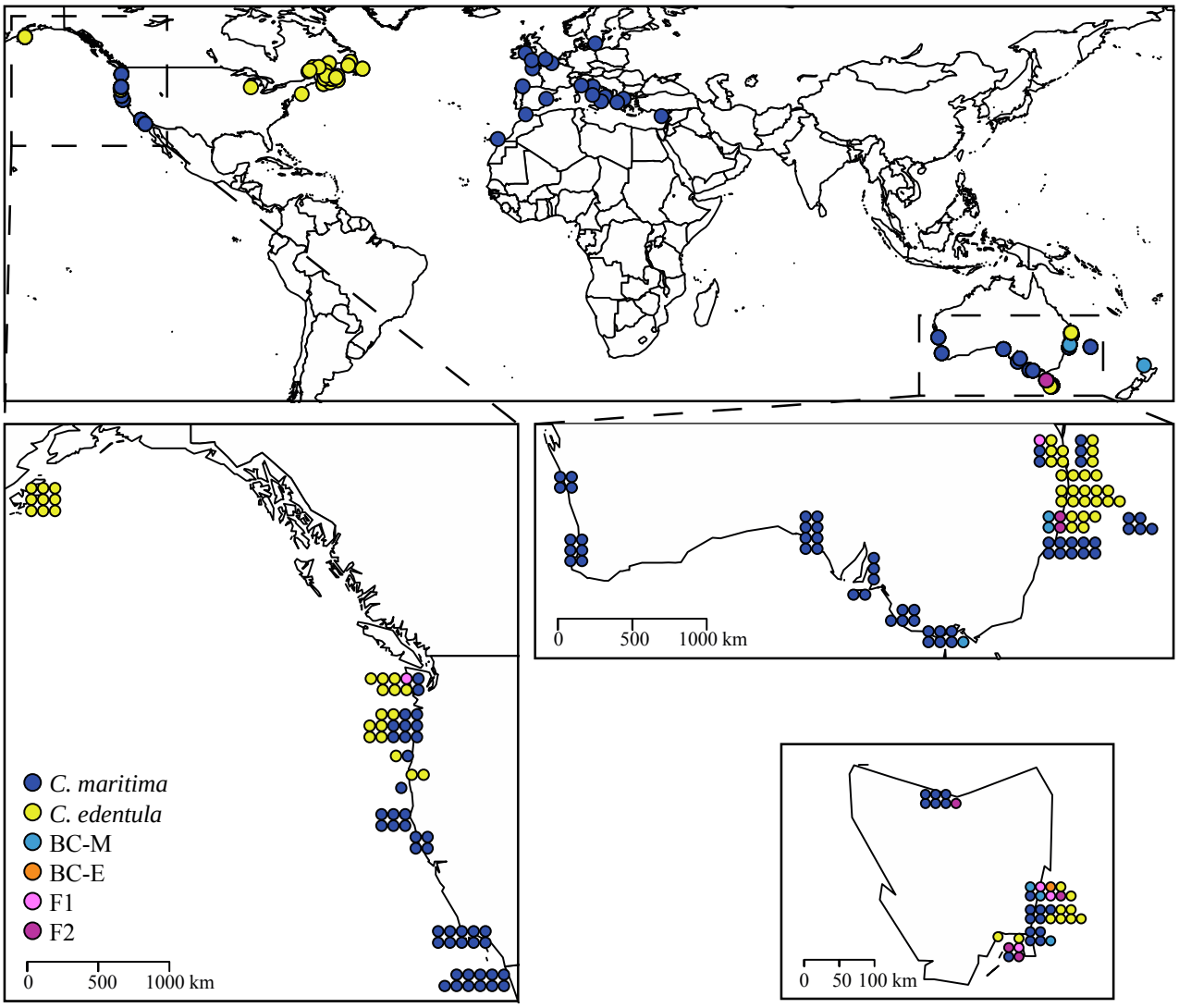
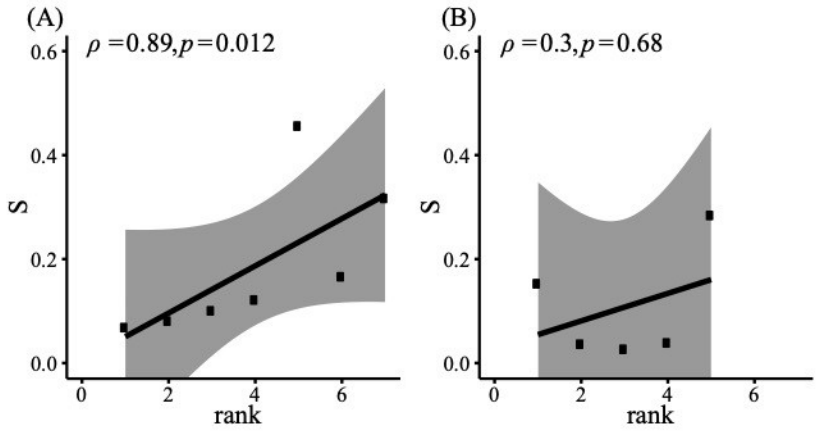


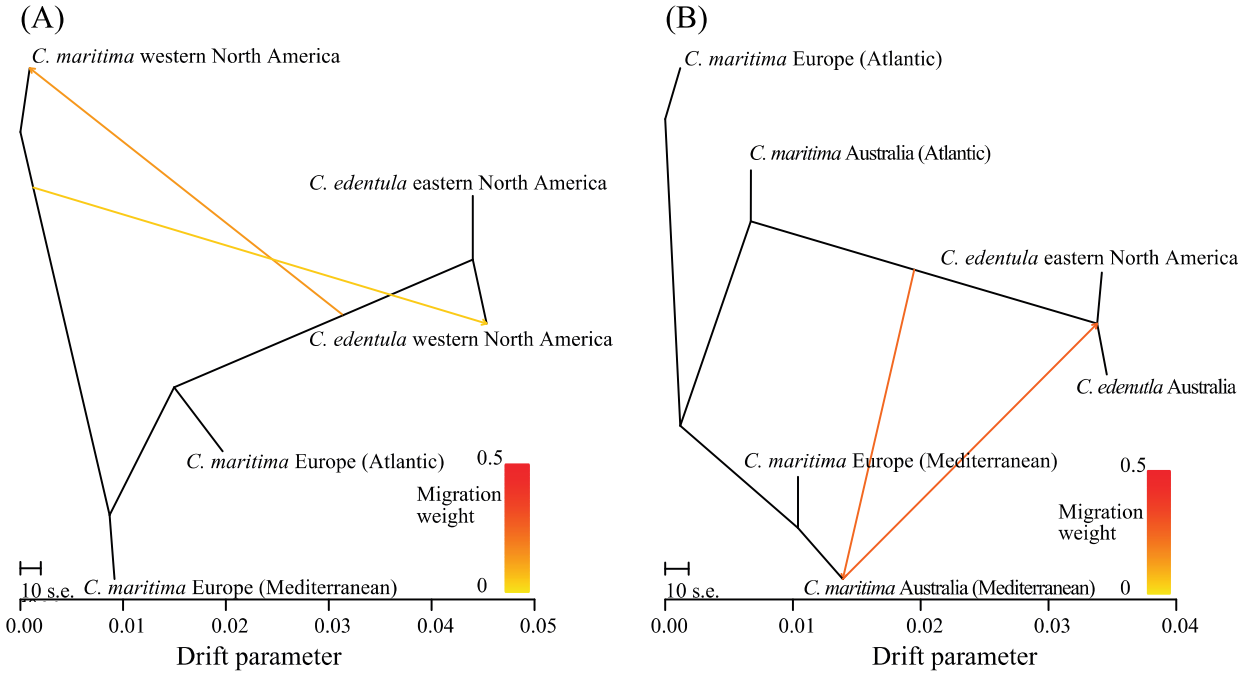
Figure 4 Geographic distribution of the hybrid assignment test by NewHybrid. Individuals are coloured according to their NewHybrid classification. A global map and close-ups of western North America, the Australian mainland and Tasmania are presented. BC-E= backcross to *C. edentula*, BC-M= backcross to *C. maritima.*



869

870 Figure 5 Results of the Spearman correlation test displayed (Table 2). The associations between
871 population mean Q values of hybrids identified using the *supervised* Admixture *run* and the ranked
872 order of populations from the first entry point of *C. maritima* (A) in eastern Australia and (B) western
873 North America.

874



875

876

877 Figure 6 Maximum likelihood trees with two migration events generated by TreeMix. Native ranges
878 and (A) western North America, (B) Australia. Individuals are grouped by species (identified
879 morphologically), likely subspecies and geographic origin.

880 Table 1 Number of individuals and sampling locations as well as mean number of individuals sampled per sampling location in each range is
881 presented.

882

Range	Phenotype	**Number of individuals**	**Number of sampling locations**	**Mean number of individuals sampled per sampling location**
Eastern North America	*C. edentula*	55	26	2.03
	C. lanceolata	2	2	1
Europe and northern Africa	*C. maritima* subsp. *integrifolia* and *baltica*	12	12	1
	C. maritima subsp. *maritima*	12	12	1
	C. maritima subsp. *islandica*	1	1	1
Western North America	*C. edentula*	39	4	4
	C. maritima	79	10	5.9
	Hybrids	2	1 (in mixed)	/
	Unknown	1	0 (in *C. edentula*)	/
	Mixed populations		3	15.6

	Total	120	17	7.05
New Zealand	Unknown	1	1	1
Australia	*C. edentula*	43	3	7.33
	C. maritima	110	11	8
	Hybrids	14	5 (in mixed)	/
	Mixed population		7	8.4
	Total	167	21	7.95

Table 2 Results of the Spearman's rank correlation test in the introduced ranges examining the association between species ancestry for *C. edentula*, *C. maritima* and hybrids or hybrids and the rank order of sampling locations based on the distance along the coastline from the first recorded case of *C. maritima* in western North America (San Francisco) or south-east mainland of Australia (Adelaide). Spearman's Rank Correlation Coefficient ρ and p values are presented for correlation between Q-value of the *supervised run* of the *C. edentula* cluster for each population in western North America and Australia and correlation between ancestry index (S) (Figure 3) and rank order of sampling locations.

			Q			**S**	
Range	Species	# populations (# individuals)	ρ	p	# populations (# individuals)	ρ	p
south-east Australia	*C. edentula*, *C. maritima*, hybrids	10 (65)	0.815	**0.004**	10 (65)	0.815	**0.004**
	Hybrids	7 (30)	0.893	**0.012**	8 (38)	0.905	**0.005**

western North America	*C. edentula*, *C. maritima*, hybrids	10 (68)	0.511	0.132	10 (68)	0.576	0.088
all sampling locations	Hybrids	5 (11)	0.300	0.683	10 (50)	0.467	0.213
western North America	*C. edentula*, *C. maritima*, hybrids	8 (47)	0.719	**0.045**	8 (47)	0.810	**0.022**
north of San Francisco	Hybrids	5 (11)	0.300	0.683	7 (30)	0.679	0.110

Table 3 Results of the f_3 statistic using TreeMix. Tests of admixture in the invasive range of Australia and western North America were done separately and both were based on three groups (hybrids, *C. edentula, C. maritima*). Hybrid classification was done according to the *supervised run* of Admixture. The f_3 statistic, the standard error of f_3 and the Z-score are reported.

Range	**Target**	**Source 1**	**Source 2**	f_3	**Standard error of f_3**	**Z-score**
Australia	Australian hybrids	Australian C. edentula	Australian C. maritima	**-0.0058**	0.0002	**-31.9723**
w. North America	w. North American hybrids	w. North America C. edentula	w. North American C. maritima	**-0.0049**	0.0002	**-23.2228**