Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

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

Hanna Rosinger[1], Armando Geraldes[2], Kristin Nurkowski[2], Paul Battlay[1], Roger Cousens[3], Loren Rieseberg[2], and Kathryn Hodgins[1]

[1]Monash University
[2]The University of British Columbia
[3]University of Melbourne School of BioSciences

June 22, 2020

Abstract

Biological invasions are accelerating, and invasive species can have large economic impacts as well as severe negative consequences for biodiversity and ecosystems. 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 a distinct source. Although morphological evidence of hybridization is generally limited to mixed populations in Australia and virtually absent elsewhere, our genetic analysis revealed relatively high levels of hybridization in Australia (34.13%), and supported the presence of hybrids in western North America (16.18%) 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.

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

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

Biological invasions are accelerating, and invasive species can have large economic impacts as well as severe negative consequences for biodiversity and ecosystems. 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 a distinct source. Although morphological evidence of hybridization is generally limited to mixed populations in Australia and virtually absent elsewhere, our genetic analysis revealed relatively high levels of hybridization in Australia (34.13%), and supported the presence of hybrids in western North America (16.18%) 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

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 (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 & Cousens, 2011; Cousens, Ades, Mesgaran, & Ohadi, 2013; Rodman, 1974, 1986) offers a

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

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 (Supplementary Information; Barbour & Rodman, 1970; Cousens et al., 2013; Rodman, 1986). 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. However, these species are closely related and cross-compatible (Rodman, 1974; Li, Cousens, & Mesgaran, 2019, Mesgaran et al., 2016). The high level of morphological variation in Australia compared to western Canada led the authors of one study (Cody & Cody, 2004) to propose the involvement of hybridization and introgression, though the mechanism of the replacement remains unclear (Barbour & Rodman, 1970; Rodman, 1986; Cousens et al., 2013).

Both *Cakile* species are found in coastal strandline habitat and spread from beach to beach via buoyant seeds (Cousens et al., 2013).*Cakile edentula* benefits from high levels of reproductive assurance, setting seeds autonomously at high rates; one of Baker's (1965) ideal weed traits. In contrast, the establishment of *C. maritima* (self-incompatible) may be initially hindered (both during initial establishment as well as subsequent range expansion) by a lack of compatible mates limiting sexual reproduction and resulting in strong Allee effects. The confirmation of hybrids in some sites in Australia led Mesgaran et al., (2016) to develop a model for the interacting species, with the novel outcome that transient hybridization could overcome Allee effects in *C. maritima* . As a consequence, we hypothesized that past hybridization with *C. edentula* could be a common feature of *C. maritima's* establishment and range expansion in western North America, Australia and New Zealand.

We used genome-wide markers derived from Genotype-by-Sequencing (GBS) to examine the invasion history of these two species in Australia and western North America and quantify the extent and distribution of hybridization. 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 the change in levels of species ancestry post-invasion in a chronosequence along the direction of invasion of *C. maritima* . We predicted that early generation hybrids are present at the leading edge of *C. maritima's* invasion into *C. edentula* -occupied areas, but later generation backcrosses with *C. maritima* are 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* .

Methods

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, Young, & Burdon , 2000). *Cakile edentula*is self-compatible and sets seed autonomously at a high rate (Barbour, 1970; Rodman, 1974). 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).

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

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 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 (Table S1, Table S2). 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.

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 (see Supplementary Information for details). Sequencing (125bp PE) was conducted on an Illumina HiSeq2500 (McGill University and Genome Quebec Innovation Centre) in two lanes.

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, & 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). The current assembly of the reference genome is found in 26,153 scaffolds with a scaffold N50 of 85,425.

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 variants that were successfully genotyped in more than 50% of individuals and removed individuals that had more than 50% missing data.

We assessed if there was a bias when mapping the reads of *C. edentula* to the reference genome of *C. maritima* (see Supplementary Information for details), but found limited evidence for such a bias (Figure S1, Figure S2). Because of this we used the BWA alignments for all downstream analyses. 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. We refer to this as the*filtered dataset* .

Genetic clustering

Population genetic structure was inferred using fastStructure (Raj, Stephens, & Pritchard, 2014) analysis. For fastStructure 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 fastStructure 10 times for k values from 2 to 10. For all fastStructure analyses we used the inbuilt function "chooseK" to obtain the best k value.

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

All following analyses were conducted in R-studio v.1.1.414 (RStudio Team, 2015) except where otherwise stated. The output of fastStructure was combined and visualized with CLUMPP (Jakobsson & Rosenberg, 2007), pophelper v.2.3.0 (Francis, 2017) and pie charts.

We failed to identify any structure in *C. edentula* in the*global thinned dataset* , even at K values higher than those the "chooseK" function recommended (Figure 1). This may have been because of the higher SNP diversity and population structuring of *C. maritima* (see Results). Consequently, to examine population structuring of *C. edentula* in the native range and the introduced range we ran two additional fastStructure analyses (10 times over the range K=2-10) (Raj et al., 2014). Here, we used the *filtered dataset*and identified polymorphic loci in the *C. edentula* native range (1912 SNPs) and thinned those SNPs to retain only one SNP per 1kb window, resulting in 705 SNPs (340 of the 705 are also in the*global thinned dataset*). The first additional fastStructure run included all eastern North American individuals (44 individuals) using those 705 SNPs and we will refer to this as *C. edentula native range dataset* . The second additional run, referred to as the *C. edentula global dataset* , included all 257 individuals using those 705 SNPs to obtain possible source populations of the introduced ranges. Finally, we ran a fourth analysis, only including polymorphic loci identified in the *C. maritima* native range, in order to obtain a clearer structuring of *C. maritima* in the native range. To obtain those SNPs, we used the *filtered dataset* , identified 14369 polymorphic loci from the native range and thinned them to retain only 1 SNP per 1kb window, resulting in 4136 SNPs. We will refer to this dataset as *C. maritima native range dataset* .

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

Genetic diversity, differentiation and inbreeding

Genetic diversity, differentiation and inbreeding 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 expected heterozygosity (H_e), observed heterozygosity (H_O), inbreeding coefficient (F_{IS}) and allelic richness (A_R). The 95% confidence intervals of F_{IS} and 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 the fastStructure Q-value assignments of the*global thinned dataset* into ten groups: (1) *C. edentula*from eastern North America (eNA_E), (2) *C. maritima* from the Mediterranean Europe (EU_Med_M), (3) *C. maritima* from the Atlantic and Baltic Europe and northern Africa (EU_At_-M), (4) Australian *C. maritima* originating from the Atlantic Europe (including the Baltic) (AUS_At_M), (5) Australian *C. maritima*originating from the Mediterranean Europe (AUS_Med_M), (6) Australian*C. edentula* (AUS_E), (7) Australian hybrids (AUS_H), (8) western North American *C. maritima* (wNA_M), (9) western North American *C. edentula* (wNA_E) and (10) western North American hybrids (wNA_H). Individuals were

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

considered hybrids if an individual had > 5 % of their genome assigned to a cluster associated with each parental species in its native range. For *C. maritima,*an individual was assigned to the Atlantic cluster (see Results; and Figure 1) when it showed > 60% assignment to this cluster; individuals were assigned to the highly admixed Mediterranean group if they showed < 60% assignment to the Atlantic cluster. To determine regional differentiation we calculated Weir and Cockerham's (1984) pairwise F_{ST} between the above ten groups using the *global thinned dataset* with VCFtools (Danecek et al., 2011).

Hybrid identification

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

(1) From the Q-scores of K=5 of the *global thinned dataset*fastStructure run (Raj et al., 2014), if an individual had > 5 % of their genome assigned to a cluster associated with each parental species in the native range.

(2) We used the program NewHybrids (Anderson & Thompson, 2002) to identify early hybrids and classify their generation. As the program is unable to deal with a large dataset, we restricted our data to 300 random SNPs that showed fixed differences between the two native ranges of the species. 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. For this package, we used all 484 SNPs which showed fixed difference between the two species. Details of the settings used are provided in the Supplementary Information.

We tested for a chronosequence correlation of the first entry point of*C. maritima* (Adelaide in Australia, San Francisco in western North America) and the extent of hybridization in *C. maritima* and hybrid individuals, with a Spearman's rank correlation test in R using the ggpubr package (Kassambara, 2020). In Australia, we only used the south- east mainland individuals. First, we tested the correlation between the Q value of the *C. edentula* cluster of the fastStructure run (*global thinned dataset*) 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 fastStructure, *C. maritima* and hybrids, or *C. edentula* , *C. maritima* and hybrids. Then, we tested the association of the S value of the HIest package and the rank order of the sampling locations along the coastline to the first entry point of *C. maritima,* using individuals that were classified as*C. maritima* and hybrid such by HIest, or *C. edentula* ,*C. maritima* and hybrids.

Results

Genetic structuring and differentiation

The fastStructure analysis of the *global thinned dataset* showed genetic structuring of *C. maritima* , *C. edentula* and hybrids. We plotted pie charts and a bar plot for K=5 (Figure 1 A, B), which was the K value that best explained the structure in the data. No genetic structure was found in the native range of *C. edentula* in contrast to the native range of *C. maritima* , which had clear geographic clustering. 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* . In fact, 43 of the 167 individuals from

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

Australia show evidence of mixed ancestry (Q-scores of > 5% from genetic clusters associated with each parental species) (Table 1). The single sample from the North Island of New Zealand included in this study was identified as hybrid (Figure 1 A, B; Table 1). In the introduced range of western North America, we identified pure *C. edentula* along with pure *C. maritima* that was genetically distinct from the native *C. maritima* clusters (Figure 1 A, B). A small number of samples (11) from Washington, Oregon and California showed evidence of hybridization.

The PCA and SPLITSTREE5 analysis confirmed the findings of fastStructure. There was clear differentiation of *C. maritima* ,*C. edentula* and hybrids in the *global thinned dataset* . The first axis of the PCA (Figure 2 A, Figure S3 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 PCA axes. 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 fastStructure) formed a monophyletic group without admixture.*C. maritima* samples were split into three groups (Figure 2 B):*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 Splitstree analysis (Figure S4) that mirrors this pattern, but provides finer *C. edentula* grouping in the native range.

The analysis using only SNPs polymorphic in *C. edentula* native range (*C. edentula native range dataset*) identified five geographically structured clusters (Figure 3 A, B), using the K value that best explained the structure in data identified by fastStructure for this reduced dataset. We assigned individuals to one group if they showed > 50 % of a cluster based on the Q value. Specifically, there was one group present each in Quebec (along the St Lawrence River), New Brunswick within the Gulf of St. Lawrence, Nova Scotia and Newfoundland, while two single samples from Lake Michigan and Rhode Island constituted the final group. The PCA (Figure 3 C) supported this grouping; the first axis explained 20.45 % of the variation and the second 13.20 %.

To identify the likely origin of introduced populations of *C. edentula,* all samples with *C. edentula* -specific polymorphic loci (*C. edentula global dataset*) were analyzed (Figure S5). The value of K that best explained structure in data was 7. The geographic clustering in the native range was similar to the previous run, although the samples from the St. Lawrence River and Newfoundland were now grouped together (Figure S5 A, B). For Australia, *C. edentula*samples grouped in the same cluster as samples from Nova Scotia.*Cakile edentula* from western North America grouped with two native range groups: samples from Washington State southward were grouped with Lake Michigan/ Rode Island samples, while in Alaska, both the Lake Michigan/ Rhode Island and the Nova Scotia cluster were apparent (Figure S5 A, B).

The *C. maritima native range dataset* identified the same two major groups for Europe (K value that best explained structure in data = 3, Figure S6 A, B) as in our *global thinned dataset* . However, unlike those results, the genetic clusters were more clearly delineated geographically, with less admixture between the Mediterranean and Atlantic regions. The PCA recapitulated this finding (Figure S6 C).

Pairwise F_{ST} (Table S3) using the *global thinned dataset* revealed clear genetic differentiation between the two-parental species originating from the native range (F_{ST}> 0.6). Within the introduced ranges the pairwise F_{ST} between the two species was still high, but generally less than in comparison to the two native ranges, perhaps reflecting some low level of ancestry of the alternate species in those individuals we classified as pure species. We used a 5% contribution of each parental species based on Q scores and clusters associated with each species in their native ranges to classify hybrids with fastStructure. Hybrids in the introduced ranges showed higher genetic differentiation from *C. edentula* than from *C. maritima*(Table S3).

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

Genetic diversity and inbreeding

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 S4); this is also true for H_e. 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 (in comparison to the native range). Hybrids of the two-species had higher H_O, H_e and A_R compared to both parental species. The inbreeding coefficient F_{IS} suggested that in its native range *C. edentula* appears slightly more inbred than *C. maritima* in its native range, although this was not significant (Table S4). The inbreeding coefficient was reduced for *C. edentula* in Australia and *C. maritima* in Australia and western North America compared to their respective native ranges; the only exception was *C. edentula* in western North America, which showed higher F_{IS} in the introduced range than in the native range. The hybrids in Australia showed similar F_{IS} values to the parental species, while the hybrids in western North America tended to have lower F_{IS} values than either parent.

Hybrid classification

The different approaches classified different proportions of individuals as hybrids. The fourteen putative hybrids included in the samples as a result of morphological identification were assigned by all analyses as hybrids, providing evidence of the accuracy of the assignments.

Classification of hybrids using FastStructure revealed 43 hybrids in Australia (34.13 %) from 12 locations, 11 hybrids in western North America (16.18 %) from five locations and one hybrid from New Zealand (Figure 1; Table 1). In western North America hybrids were found in each of two locations in California and Oregon and in one location in Washington.

NewHybrids analysis revealed 26 hybrids (Table 1) with 22 hybrids in Australia (17.46 %), three in western North America (4.11 %) and one in New Zealand. All NewHybrids hybrids were concordant with the 55 hybrids detected by the fastStructure analysis. In Australia, F1 and F2 hybrids were detected in the current sympatric zones where individuals of both species morphologies were clearly identifiable in the populations. Hybrids (Figure S3 B) grouped in the PCA according to their generation, as F1 and F2 hybrids grouped in the middle of the two species, whereas backcrosses grouped closer to species they backcrossed to. In this same PCA the additional hybrids identified with fastStructure grouped with *C. maritima,* suggestive of further backcrossing to that species.

HIest (Fitzpatrick, 2012) classified 74 individuals as hybrids (Figure 4 A, B; Table 1, Table S5), 68 hybrids in Australia (53.97%, 15 F2, five BC-E, 48 BC-M), five hybrids in western North America (7.35 %, one F2, one BC-E, two BC-M) and one hybrid in New Zealand (one BC-M). All individuals identified by NewHybrids were also identified by HIest as hybrids, although the hybrid type varied slightly between the two analyses. Australian hybrids were from 14 locations. Backcrosses *to C. maritima* were identified in this analysis in regions where *C. maritima* has not been recorded for many decades. In the current sympatric zone (New South Wales, Queensland and Tasmania) this analysis identified F2 hybrids as well backcrosses to both parental species. In western North America hybrids were detected in three locations. One F2 hybrid was identified in California, one BC-M was identified in Oregon and in Washington one each of F2, BC-E and BC-M (Figure 4 B) were identified.

Ancestry assignment to each species (using S values) of populations was correlated with the distance from the origin of *C. maritima* in both south-eastern Australia (excluding Tasmania) and western North America (Table 2). In Australia, this pattern was significant when testing across all samples (R=-0.77, $p < 0.01$) and when testing within *C. maritima* and their hybrids (R=-0.77, $p < 0.05$) or the hybrids (identified by HIest) alone (R==-0.82, $p<0.05$). However, in western North American it was only significant when testing across all samples (R=-0.79, $p < 0.05$), although the patterns of species ancestry were all in the same direction.

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

Similar patterns were observed when using Q values from fastStructure to assign the proportion of the genomes to each species, although the hybrid only analysis was not significant.

Discussion

Our aim was to identify probable source regions (from Europe and eastern North America) for the invasions (in Australia and western North America) and to assess patterns of hybridization between *C. edentula* and *C. maritima* in these two introduced ranges. In both regions, a similar pattern of invasion and replacement has been identified using historical records (Barbour & Rodman 1970, Rodman 1986, Cousens et al., 2013). There have been several previous studies examining the population genetic structure of *C. edentula* and*C. maritima* in their native ranges (Europe (Clausing, Vickers, Kadereit, 2000; Kadereit, Arafeh, Somogyi, & Westberg, 2005; Westberg, 2005), Africa (Gandour, Hessini, & Abdelly, 2008), eastern and western North America (Gormally & Donovan, 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.

Our analysis provides evidence that *C. edentula* populations in Australia were likely sourced from the Nova Scotia region, while in western North America *C. 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 divergence of the western North American *C. maritima* populations from the other *C. maritima* samples suggests an un-sampled source, although they show some affinity to samples from the Mediterranean. Importantly, we found frequent hybridization in Australia (34.13 %) as well as the first genetic evidence of hybrids in western North America (16.18 %) 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 in contact and new hybrids are still being formed, it would be difficult to determine visually that hybridization has ever taken place, since backcrossing soon hides its 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 highly 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.

Native range patterns

Our analysis provided evidence of geographic structuring in the *C. edentula* native range, at a much finer grain than currently recognized taxonomically (Figure 3). Samples from Quebec, Newfoundland, Nova Scotia and New Brunswick fall into separate clusters, likely within *C. edentula* subsp. *edentula* var. *edentula* as this subspecies is the only one described in this region of the North American Atlantic coast (Rodman, 1974). Divergence between the two single samples from Lake Michigan and Rhode Island drive the pattern along EV1 in the PCA analysis, explaining 20 % of the variation. They also grouped together in one cluster of the Structure analysis; those samples might belong to the Atlantic coast variety of *C. edentula* subsp. *edentula*var. *edentula* as it is known to have invaded Lake Michigan in historical times (Rodman, 1974; Huebner, 2009), where it now coexists with the Great Lakes endemic var. *lacustris* . A second possibility, suggested by Gormally and Donovan (2011), but without morphological evidence, is that var. *lacustris* has dispersed to the Atlantic. Genetically distinct regional variation is not surprising, as the directions of currents

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

and the influences of geological features on seed dispersal can be highly predictable (Lapointe, 2000). Similar conclusions have been reached in the Mediterranean by Westberg (2005) and Gandour et al. (2008). *Cakile edentula* subsp. *harperi*occurs in areas south of the populations sampled in our study (Rodman, 1974), but comprehensive studies of herbarium samples by Rodman (1974) and Cousens et al., (2013) have found no morphological evidence that subsp. *harperi* has been introduced anywhere outside its native range.

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

Cakile edentula showed lower genetic diversity than *C. maritima* in their native ranges as measured by allelic richness and expected heterozygosity (Table S4) and showed less variation along the PCA axes and in the SPLITSTREE network analysis (Figure 3). 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). In addition, it has been suggested that *C. edentula* is a relatively recently derived species that may have originated from a long-distance dispersal event from *C. maritima* (Rodman, 1974). It is possible that such a bottleneck, in combination with frequent self-pollination and an ephemeral life history in a colonizing annual species, would serve to reduce genetic diversity in *C. edentula* compared to *C. maritima* .

Introduced range patterns

Australia and New Zealand

Pure *C. edentula* populations still remain in Australia (Queensland, New South Wales and Tasmania) and our analyses show that they likely originate from populations located in Nova Scotia.*Cakile edentula* allelic richness and H_O did not change considerably in Australia compared to the native range (Table S4), which is inconsistent with a strong invasion bottleneck. Indeed, the negative F_{IS} value indicates an excess of heterozygotes compared to expectations based on allele frequencies, although the confidence interval overlaps with zero. It would be interesting to determine if this reduction in F_{IS} is associated with undetected low-level *C. maritima* ancestry.

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 (Rodman, 1976, 1986; Cousens et al., 2013; Ohadi et al., 2016). 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). 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 allelic richness and H_O values than its native range, consistent with admixture of multiple source populations and/or hybridization with *C. edentula.*

Our data provides substantial evidence for extensive hybridization in Australia between the two species. As expected, Australian hybrids had higher genetic diversity than both parental species. Furthermore, the pattern of hybrid ancestry was geographically structured and reflected the historical invasion route of *C. maritima* in south-eastern Australia (Table 2; Figure 1, Figure 4). NewHybrids and HIest confirmed the presence of early generation hybrids where both species still co-occur. Some mixed populations in Australia show pure genotypes of both parental species and early generation hybrids, demonstrating on-going hybridization of the two taxa. In areas where*C. edentula* still persists, backcrossing to *C. edentula*has also occurred, but is rare, and recent backcrosses to *C. maritima* appear to be more common. In those parts of Australia

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

where *C. maritima* has already appeared to have replaced *C. edentula* (i.e. where no *C. edentula* phenotypes remain; Rodman, 1986; Cousens et al., 2013), evidence is consistent with past hybridization between the species and repeated backcrossing to *C. maritima* (Figure 1, Figure 4). In areas of Western Australia, where *C. edentula* has never been confirmed, evidence of hybridization with *C. edentula* was also identified, confirming a previous observation by Ohadi et al., (2016). The sample from New Zealand was identified as a recent hybrid (BC-M) where the same replacement of *C. edentula* by *C. maritima* has also taken place (Cousens & Cousens, 2011).

Western North America

Our results revealed that *C. edentula* in western North America most likely originated from two sources in eastern North America. By contrast, we found that western North American *C. maritima* is genetically distinct from any native range *C. maritima* sampled. One possibility is that a bottleneck resulted in substantial divergence from an existing sampled source. Alternatively, the western North American *C. maritima* may have originated from a yet un-sampled source. Further sampling in the native range of *C. maritima* is needed to fully distinguish between these hypotheses. However, our fastStructure analysis (Figure 1) was consistent with an introduction from a cluster sampled in the Mediterranean region, and pairwise F_{ST} (Table S3) values also show greater similarity between the western North American *C. maritima* and the "Mediterranean" group than with the "Atlantic" group. *Cakile edentula* and *C. maritima* in western North America showed, as in Australia, no reduction H_O and A_R, which may reflect the impacts of undetected hybridization, large founding populations, or multiple introductions.

We identified 11 hybrid samples from five locations in western North America; of these, three samples from two populations were of recent origin. Specimens of hybrids based on morphological identification are largely unknown for this region, either in herbaria or in the field (Rodman 1974). Although the fitness and demographic consequences of 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. In addition, the mechanism driving differences in hybridization rates in western North America compared to Australia is unclear.

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. Rodman, 1974; Mesgaran et al., 2016; Li et al., 2019) 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. 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 (Li et al., 2019), which follows the 'SI × SC rule' of unilateral incompatibility (Harrison & Darby, 1955; Pickup et al., 2019). Li, Mesgaran, Ades and Cousens (2020) have shown that early generation backcrosses between *C. maritima* and hybrids had traits more consistent with outcrossing (e.g. floral display, greater flower production) and outperform other hybrid crosses.

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

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

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. 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 and Rieseberg, 2006). Indeed, Cody and Cody (2004) proposed the intriguing possibility of adaptive introgression in this system. 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.

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.

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.

Funding

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

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.

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

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.

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

References

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.

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.

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.

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, D. G., Caseys, C., Cousens, R. D., Hahn, M. A., Heredia, S. M., Hubner, S., ... Rieseberg, L. H. (2015). What we still don't know about invasion genetics. *Molecular ecology* , *24* (9), 2277-2297.

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.

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.

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.

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

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.

Fitzpatrick, B. M. (2012). Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evolutionary Biology* , *12* (1), 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.

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.

Gormally, C. 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.

Harris, K., & Nielsen, R. (2016). The genetic cost of Neanderthal introgression. *Genetics* , *203* (2), 881-891.

Harrison, B. J., & Darby, L. A. (1955). Unilateral hybridization. *Nature* , *176* (4490), 982-982.

Hoffmann, B. D., & Broadhurst, L. M. (2016). The economic cost of managing invasive species in Australia. *NeoBiota* , *31* , 1.

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.

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.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* , *23* (14), 1801-1806.

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.

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.

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., & Prodohl, P. A. (2013). DiveRsity: An R package for the estimation of population genetics parameters and their associated errors. Methods Ecol Evol. 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, Quebec (Canada). *Marine Micropaleontology* , *40* (1-2), 43-65.

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

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.

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.

Li, H. & Durbin, R.. (2009). Fast and accurate short read alignment with Burrows- Wheeler transform. *Bioinformatics* 25, 1754–1760.

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.

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.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular ecology* , *25* (11), 2337-2360.

Pickup, M., Brandvain, Y., Fraisse, 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.

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.

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.

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* , *197* (2), 573-589.

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.

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.

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

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.

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. *Elem Sci Anth* , *1* , p.000008. doi: http://doi.org/10.12952/journal.elementa.000008

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.

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hubner, S., ... Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary applications* , *9* (7), 892-908.

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.

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.

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.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. Evolution, 38, 1358–1370.

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

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.

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.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., Francois, R., ... Kuhn, M. (2019). Welcome to the Tidyverse. *Journal of Open Source Software* , *4* (43), 1686.

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.

Figures and figure captions

Hosted file

`image1.emf` available at https://authorea.com/users/335639/articles/461485-the-tip-of-the-iceberg-genome-wide-marker-analysis-reveals-hidden-hybridization-during-invasion

Figure 1. FastStructure results of the *global thinned dataset* . (A) A Disrupt plot for K=5. Individuals are ordered according to their cluster association. 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=5, fastStructure 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 Disrupt plot.

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

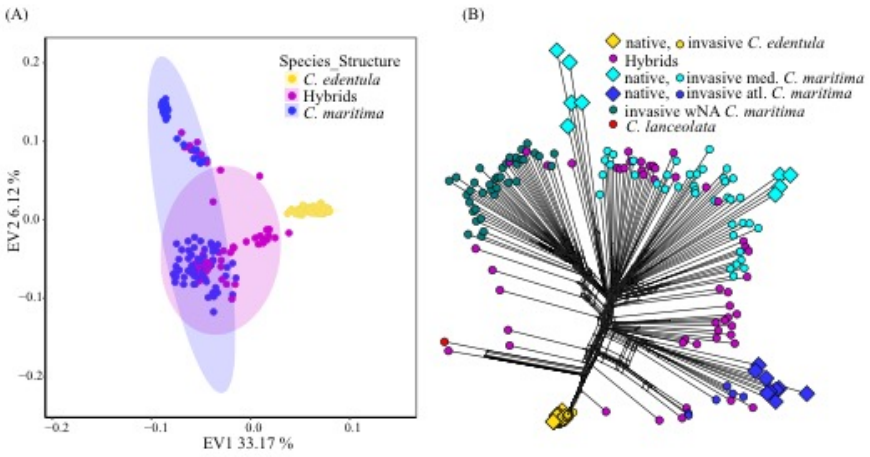


Figure 2 (A) Principal component analysis of the *global thinned dataset* . First two eigenvectors are presented. Individuals are coloured according to their species and hybrid status based on the fastStructure results of K=5. Ellipses indicate the 95 % confidence range of the cluster. (B) Splitstree network of the *global Splitstree dataset* . Individuals are coloured according to their predominant fastStructure cluster (K=5 of the *global thinned dataset*). The shapes indicate native vs. invasive range.

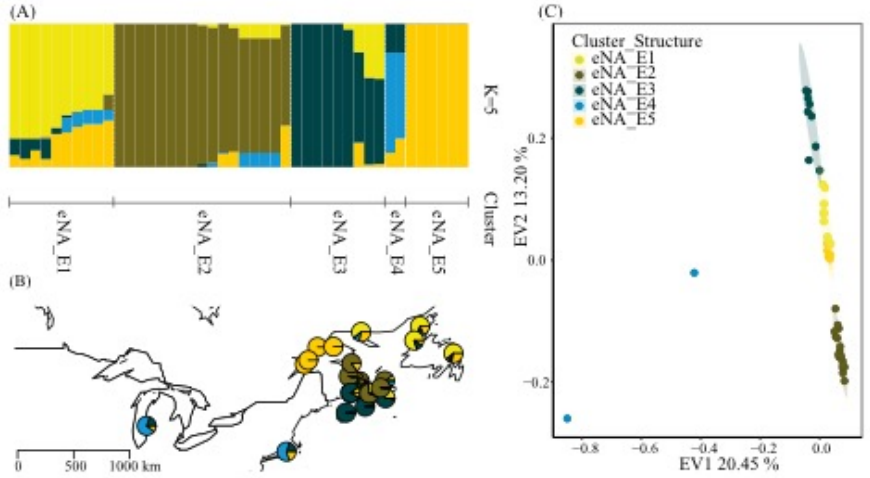


Figure 3 Visualization of the *C. edentula* native *range dataset* fastStructure results. (A) Disrupt plot of fastStructure results for K=5. Individuals are ordered according to their cluster association, and clusters are labelled as eastern North America *C. edentula* cluster 1-5 (=eNA_E1-5). (B) Map of eastern North American sampling locations with fastStructure proportions (K=5) per population indicated by colour pie charts. (C) Principle Component Analysis display of the first two Eigenvectors. Individuals were grouped and coloured according to their fastStructure cluster (eNA_E1-5). Ellipses indicate the 95 % confidence range of the cluster.

Hosted file

`image4.emf` available at https://authorea.com/users/335639/articles/461485-the-tip-of-the-iceberg-genome-wide-marker-analysis-reveals-hidden-hybridization-during-invasion

Figure 4 Results of a hybridization assignment test implemented by HIest. (A) Association of ancestry

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

index (S) and interclass heterozygosity (H) are given for western North America (left) and Australia (right). Individuals are coloured according to their HIest classification. BC-E= backcross to *C. edentula* , BC-M= backcross to *C. maritima* , E= *C. edentula,* F2, M= *C. maritima.*(B) The geographic distribution of individuals classified hybrids by HIest, coloured as in (A). A global map and close-ups of western North America, the Australian mainland and Tasmania are presented.

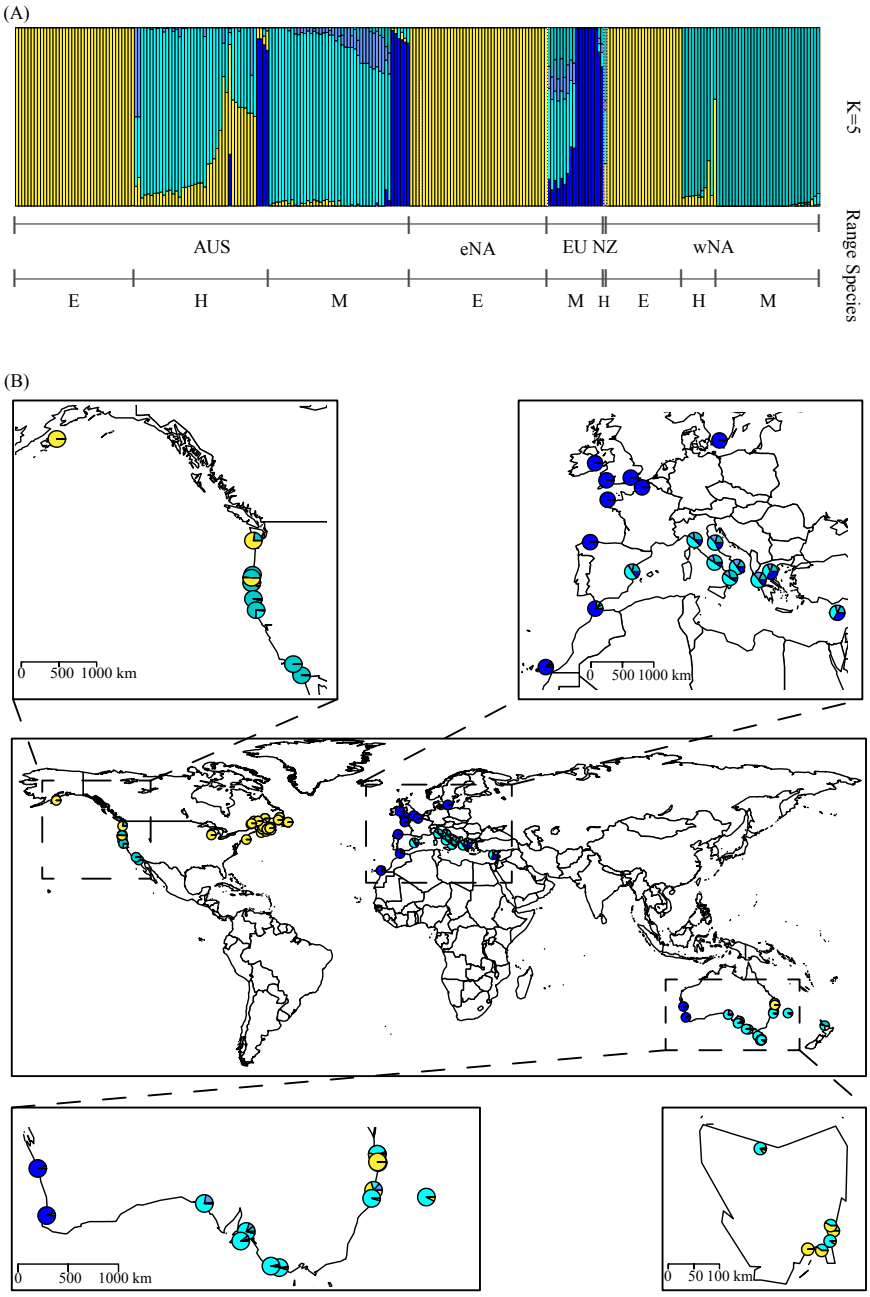
Table 1 FastStructure, NewHybrids and HIest classification of hybrid ancestry for *Cakile* individuals sampled in Australia, western North America, New Zealand, eastern North America, Europe and northern Africa. The number classified as a pure species or to a hybrid generation (BC-E= backcross to *C. edentula* , BC-M= backcross to*C. maritima*) is shown and percentage per range is given. Note that fastStructure does not identify the hybrid class.

Program	Range	*C. edentula*	*C. maritima*	F1	F2	BC-E	BC-M
FastStructure	Australia	38 (30.16%)	45 (35.71 %)				
	western North America	24 (35.29 %)	33 (48.53 %)				
	New Zealand						
	eastern North America	44 (100 %)					
	Europe and northern Africa		18 (100 %)				
Total							
NewHybrids	Australia	38 (30.16%)	66 (52.38 %)	4(3.17 %)	8(6.35 %)	1(0.79 %)	9(7.14
	western North America	24 (35.29 %)	41(60.29 %)	1(1.47 %)			2(2.94
	New Zealand						1 (100
	eastern North America	44 (100 %)					
	Europe and northern Africa		18 (100 %)				
Total				5	8	1	11
HIest	Australia	38 (30.16%)	20(15.87 %)		15(11.90 %)	5(3.97 %)	48(38.1
	western North America	24 (35.29 %)	39(57.35 %)		2(2.94 %)	1(1.47 %)	2(2.94
	New Zealand				1 (100 %)		
	eastern North America	44 (100 %)					
	Europe and northern Africa		18 (100 %)				
Total					18	6	50

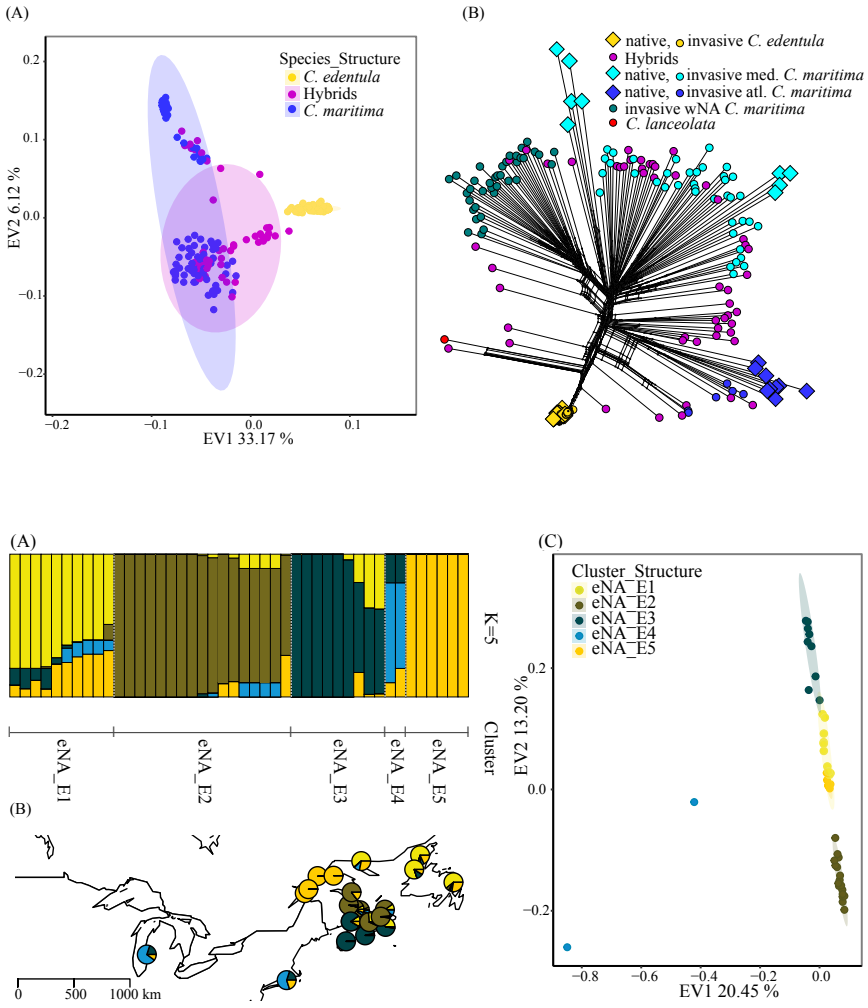
Table 2 Results of the Spearman's rank correlation test in the introduced ranges examining the association between species ancestry for*C. edentula* , *C. maritima* and hybrids; *C. maritima*and hybrids; hybrids and the rank order of sampling locations based on the distance along the coastline from the first recorded case of*C. maritima* in western North America (San Francisco) or south-east mainland of Australia (Adelaide). Spearman's Rank Correlation Coefficient R and p values are presented for correlation between Q-value (Figure 1) of the *C. edentula* cluster for each population in western North America and Australia and correlation between ancestry index (S) (Figure 4) and rank order of sampling locations.

Range	Species	Q	Q	S	S
		p	R	p	R
south-east Australia	*C. edentula*, *C. maritima*, hybrids	**0.003**	0.778	**0.004**	-0.767
	C. maritima, hybrids	**0.003**	0.833	**0.014**	-0.770
	Hybrids fastStructure	0.242	0.600		
	Hybrids HIest			**0.034**	-0.821
western North America	*C. edentula, C. maritima*, hybrids	**0.006**	0.830	**0.010**	-0.794
	C. maritima, hybrids	0.216	0.500	0.389	-0.357
	Hybrids fastStructure	0.517	0.400		
	Hybrids HIest			1.000	-0.500

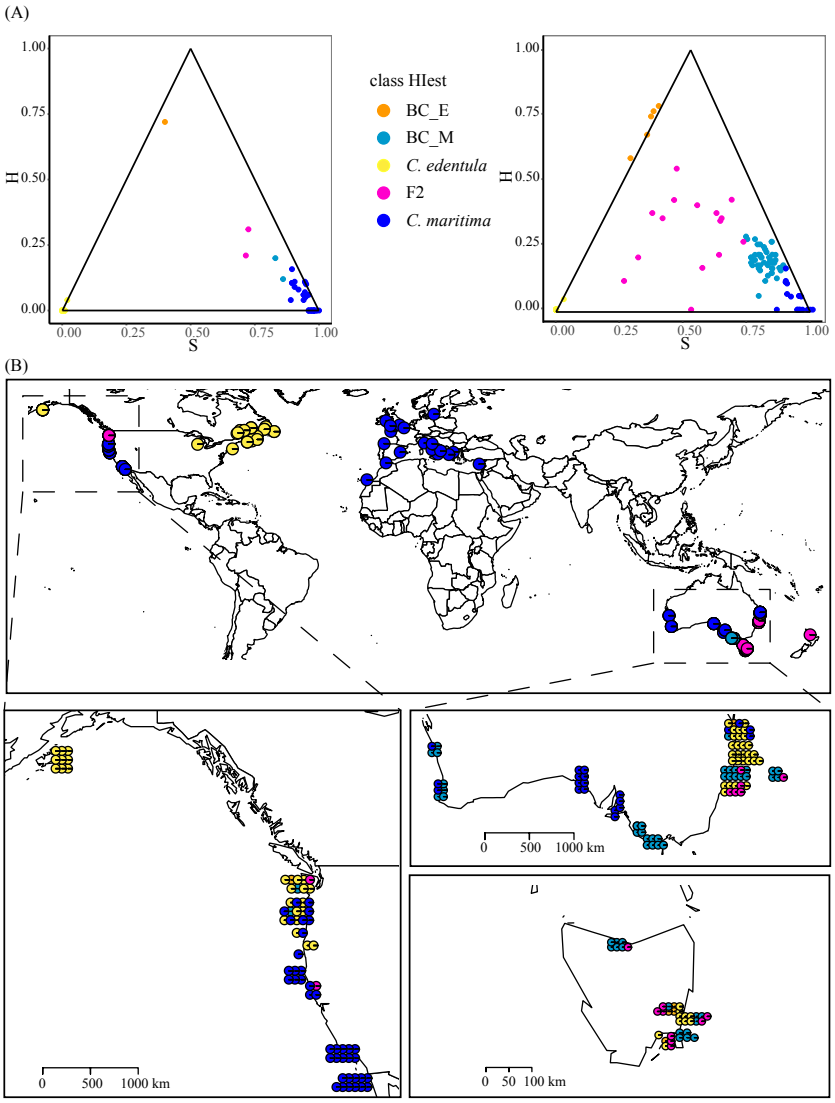
Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.



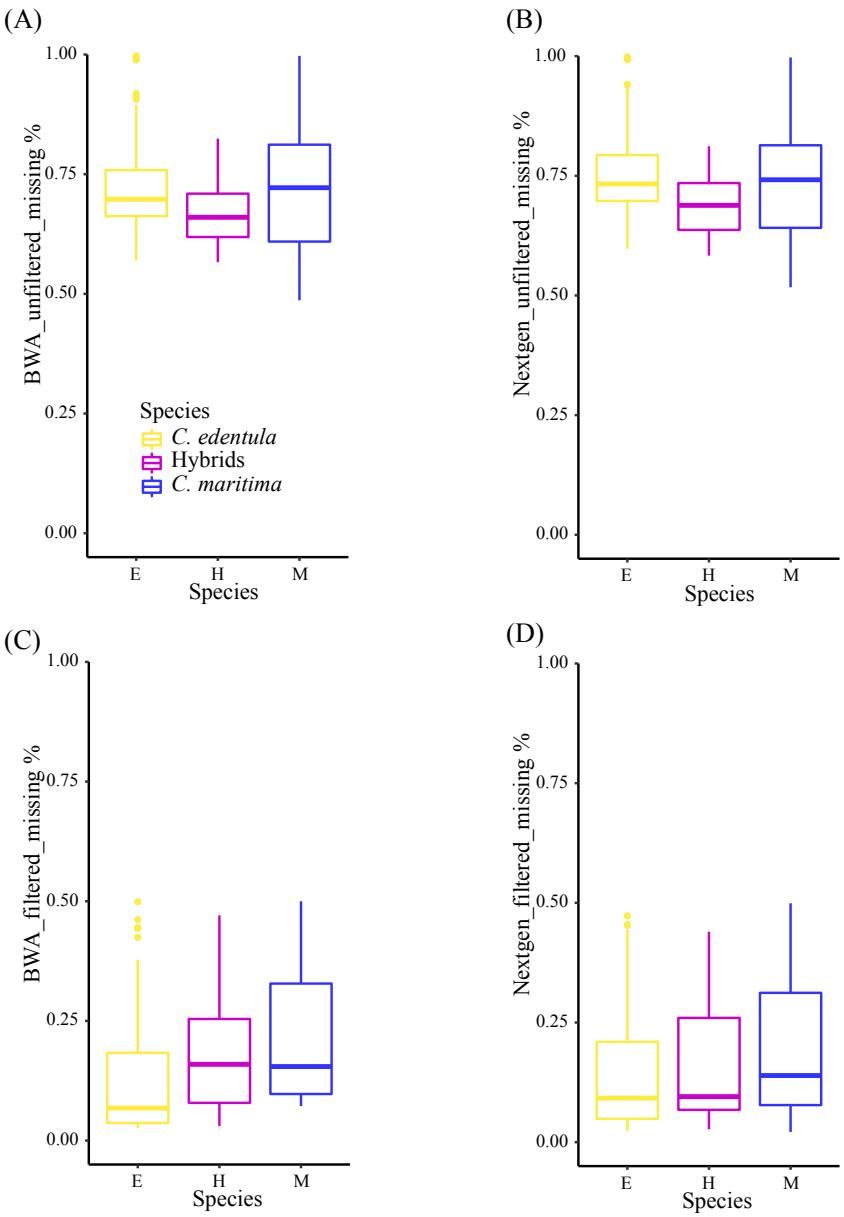
Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.



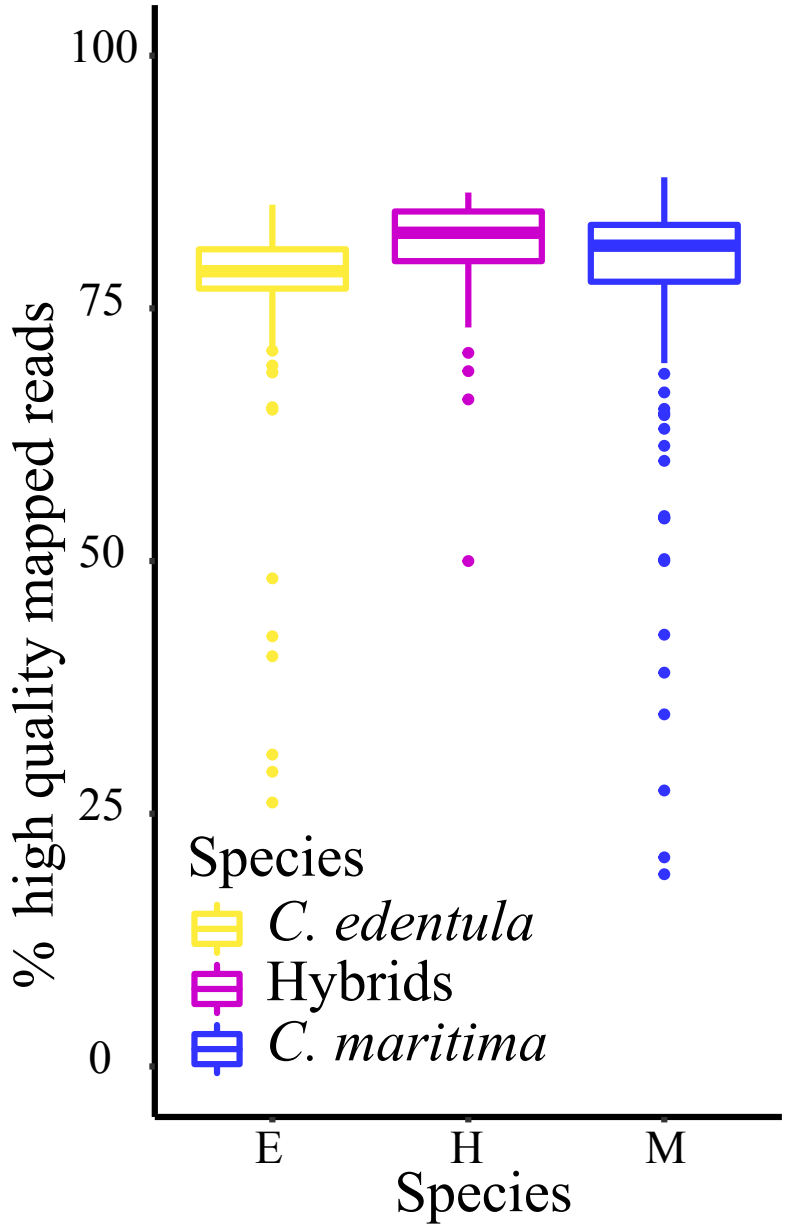
Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.



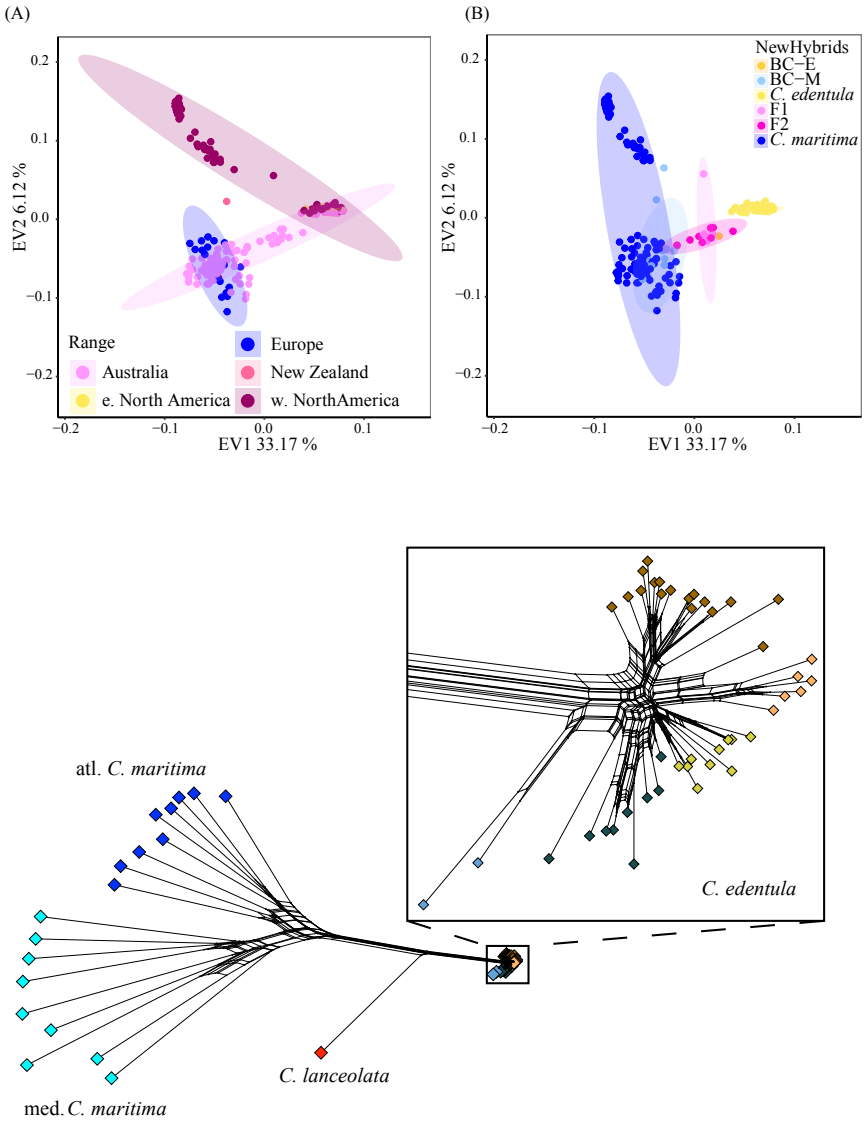
Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.



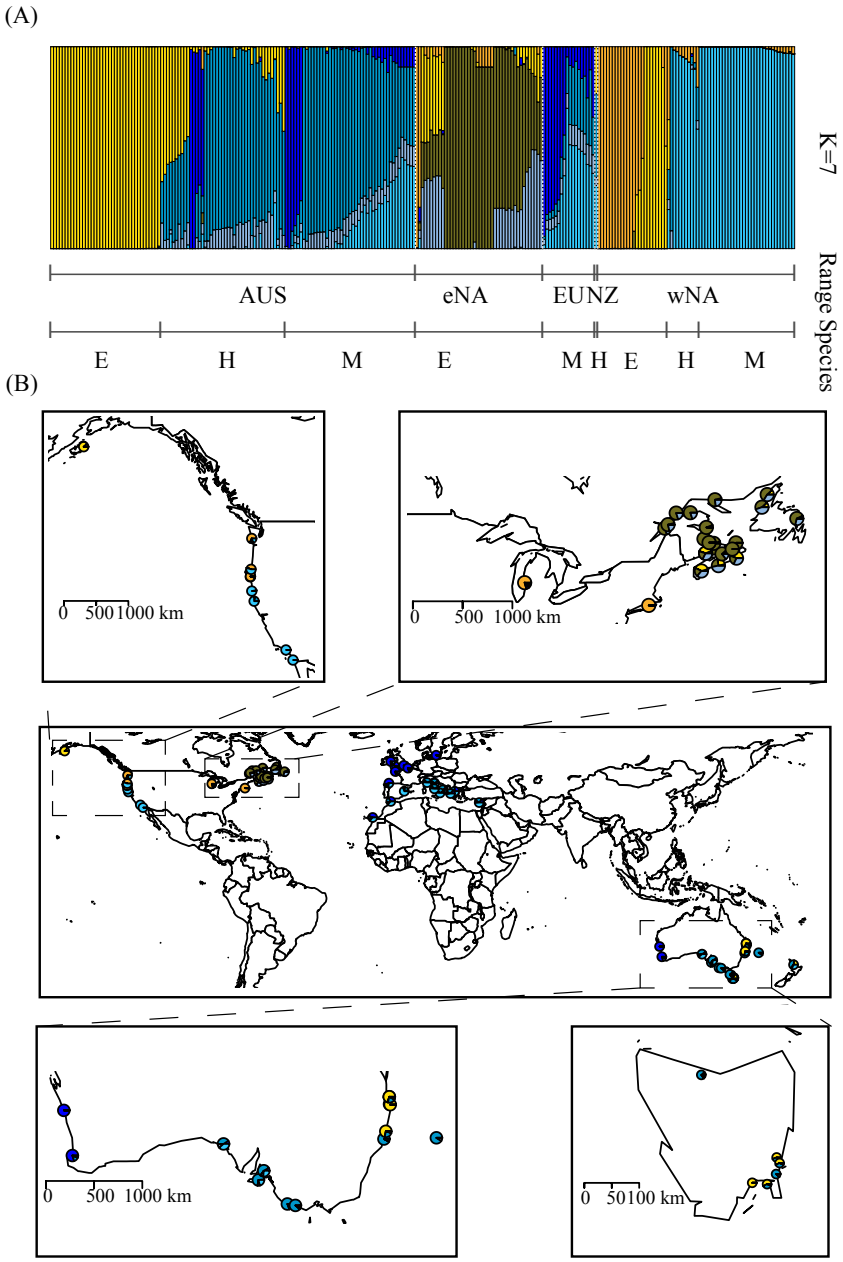
Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.



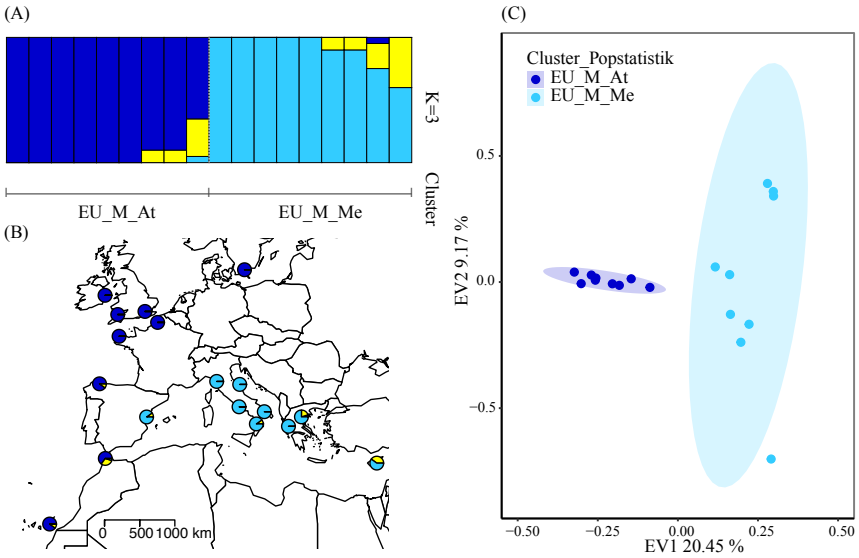
Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.



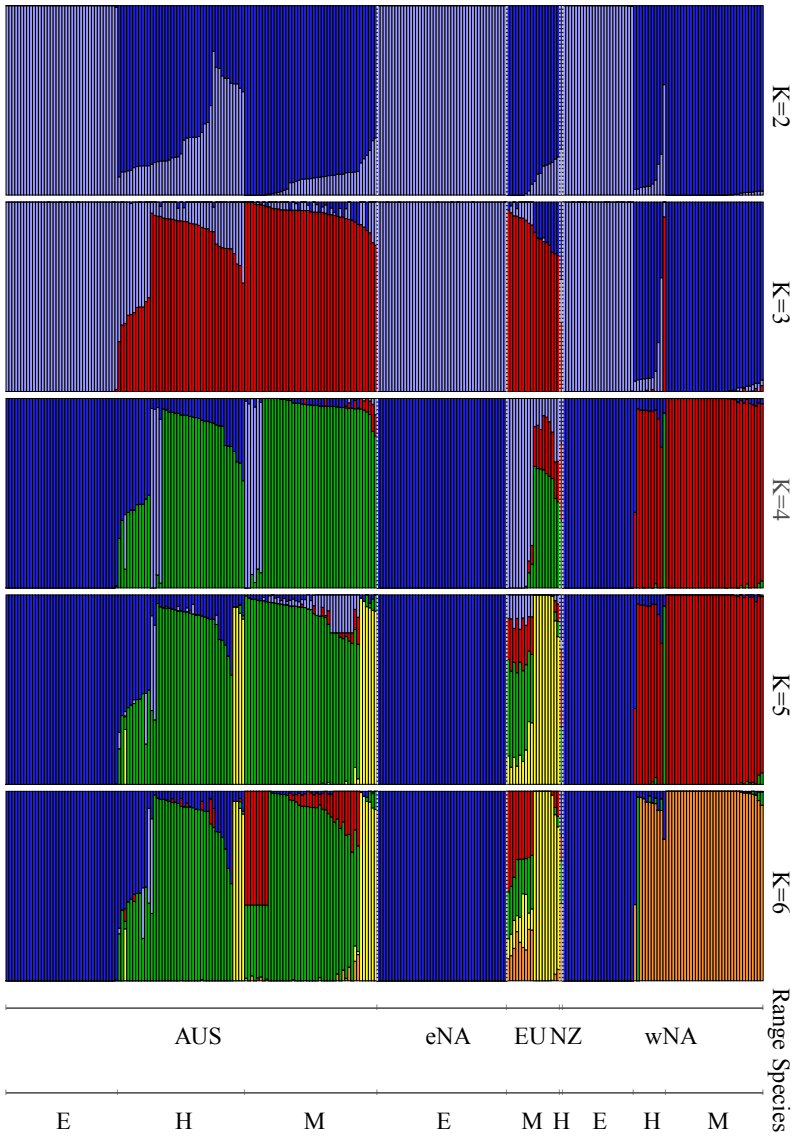
Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.



Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

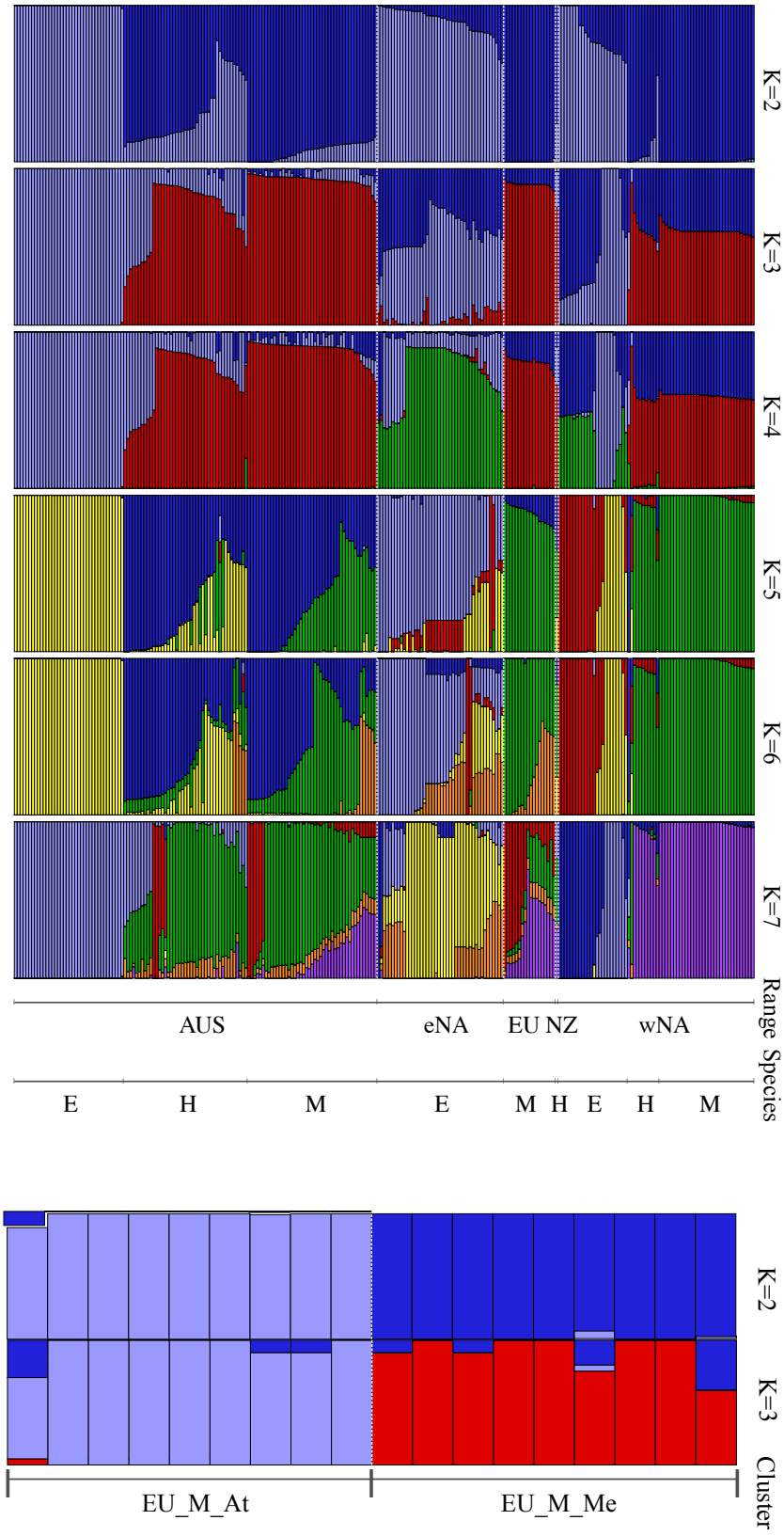


Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.



Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.



Posted on Authorea 22 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159284861.17694590 — This a preprint and has not been peer reviewed. Data may be preliminary.

Hosted file

`Table S2.xlsx` available at https://authorea.com/users/335639/articles/461485-the-tip-of-the-iceberg-genome-wide-marker-analysis-reveals-hidden-hybridization-during-invasion