NATURAL CHIMERISM IN THE INVASIVE COLONIAL ASCIDIAN DIDEMNUM VEXILLUM

A Thesis submitted to the faculty of
San Francisco State University
In partial fulfillment of
the requirements for
the Degree

Master of Science

In

Biology: Marine Biology

by

Rachel Beth Weinberg

San Francisco, California

August 2019

CERTIFICATION OF APPROVAL

I certify that I have read *Natural chimerism in the invasive colonial ascidian* Didemnum vexillum by Rachel Beth Weinberg, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirement for the degree Master of Science in Biology: Marine Biology at San Francisco State University.

C. Sarah Cohen, Ph.D.
Professor

Colleen Ingram, Ph.D.
Adjunct Professor

Eric Routman, Ph.D.
Professor

NATURAL CHIMERISM IN THE INVASIVE COLONIAL ASCIDIAN DIDEMNUM VEXILUM

Rachel Beth Weinberg
San Francisco, California
2019

Colonial, asexually reproducing organisms are typically treated as genetically homogenous entities composed of clonal units. However, processes such as fusion and somatic mutation may lead to the integration of multiple genetic lines in a single physiologically continuous organism. This state, known as chimerism, has been documented in a broad range of sessile modular taxa, including plants, fungi, slime molds, and colonial metazoans such as ascidians. In the invasive colonial ascidian *Didemnum vexillum,* naturally occurring chimerism has been documented from the presence of more than two alleles at microsatellite loci, however factors such as homozygosity and spatial separation of genets within a colony may mask chimerism using this detection method alone. In order to more accurately quantify the prevalence of chimerism in introduced *D. vexillum* populations, zooids were dissected from 35 colonies and individually genotyped. Both spatial and molecular masking of chimerism was found in a proportion of the chimeric colonies, indicating that the sampling strategy applied here has the potential to identify chimerism in *D. vexillum* colonies that appear monogeneic from single tissue samples. In Umpqua, Oregon, 70% of the colonies sampled were chimeric, which is the highest prevalence of chimerism reported in any *D. vexillum* population studied to date. In Half Moon Bay, 30% of the samples were chimeric, which is similar to the rates of chimerism previously estimated in this population using single tissue samples. The significant differences in chimerism between the two populations were not associated with significant differences in genetic diversity or inbreeding, as measured by expected and observed heterozygosity. Relatedness was significantly higher among genets within chimeric colonies than across genets in the

population as a whole, providing evidence from two natural populations that chimerism in *D. vexillum* is genetically mediated. These results show that there may be extreme variation in chimerism prevalence across populations, providing a basis for future studies to investigate the causes of this variation and its potential role in facilitating the successful establishment of non-native species.

I certify that the Abstract is a correct representation of the content of this thesis.

Chair, Thesis Committee

8-6-19

Date

PREFACE AND/OR ACKNOWLEDGEMENTS

I am grateful for the support of my family, partner, and close friends, whose encouragement, understanding, and patience was, and is, an unfailing source of inspiration and motivation. I am also very thankful for the faculty at SFSU who taught and mentored me, as well as the graduate and undergraduate students who volunteered their time to help with this project. This work is dedicated to my father, Leonard Weinberg, whose love of tropical fish inspired my passion for marine biology and who supported my studies even as he battled with his own illness.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figures Page

LIST OF APPENDICES

Appendix Page

1. Introduction

Intra-organismal genetic heterogeneity, or chimerism, occurs broadly across the tree of life. It has been documented in a variety of colonial and modular taxa ranging from ascidians (Bancroft, 1903; Oka & Watanabe, 1957; Westerman, Dijkstra, & Harris, 2009), cnidarians (Cadavid, Powell, Nicotra, Moreno, & Buss, 2004; Chang, Orive, & Cartwright, 2018; Hart & Grosberg, 1999; Maier, Buckenmaier, Tollrian, & Nürnberger, 2012; Puill-Stephan, van Oppen, Pichavant-Rafini, & Willis, 2012; Schweinsberg, Weiss, Striewski, Tollrian, & Lampert, 2015; Shenk, 1991) and sponges (Blanquer & Uriz, 2011; Gauthier & Degnan, 2008; Ilan & Loya, 1990) to plants (Gill, Chao, Perkins, & Wolf, 2003; Mario Pineda-Krch & Poore, 2004), algae (Poore & Fagerström, 2000; Santelices, González, Beltrán, & Flores, 2017), and aggregating slime molds (Castillo, Queller, & Strassmann, 2011; Foster, Fortunato, Strassmann, & Queller, 2002). The acquisition and maintenance of chimerism has been extensively studied as an adaptive mechanism in plants (Frank & Chitwood, 2016; Gill et al., 2003) and macroalgae (Monro and Poore 2004; Santelices et al. 2017), as genetic variation within a physiologically integrated organism provides a substrate for selection to act upon during a single generation (Sommerfeldt, et al. 2003; Pineda-Krch and Lehtilä 2004; Santelices et al. 2017). A significant amount of experimental work has demonstrated that many modular invertebrates undergo allogeneic fusion to form chimeric colonies (Bancroft, 1903; Grosberg & Quinn, 1988; Hart & Grosberg, 1999; Nawrocki & Cartwright, 2012;

Rosengarten & Nicotra, 2011; Saito, Hirose, & Watanabe, 1994), however some uncertainty remains regarding the function and frequency of chimerism under natural conditions. Recent evidence suggests that natural chimerism may be common in colonial invertebrates such as stony corals (Becheler et al., 2017; Maier et al., 2012; Puill-Stephan et al., 2012; Schweinsberg et al., 2015), sponges (Blanquer & Uriz, 2011), and ascidians (Sommerfeldt, Bishop, and Wood 2003; Westerman, Dijkstra, and Harris 2009; Smith et al. 2012; Clancy 2015).

Chimerism has been hypothesized to contribute to the success of invasive species by providing a reservoir for genetic diversity (Crawford & Whitney, 2010; M. Pineda-Krch & Lehtilä, 2004) and providing the chimeric organism with a greater range of environmental tolerances (Rachel Ben-Shlomo, 2017; Rinkevich, 2004), advantages that may reduce some of the barriers to establishment faced by nonnative species. Fusion and chimerism have been documented in both the native and nonnative range of the invasive colonial ascidian *Didemnum vexillum* (Clancy 2015; Smith et al. 2012; Watts, Hopkins, and Goldstien 2019). Native to Japan (Bullard et al. 2007; Stefaniak et al. 2009; Lambert 2009), *D. vexillum* can be found throughout Europe (Gröner, Lenz, Wahl, & Jenkins, 2011; Ordóñez et al., 2015; Osman & Whitlatch, 2007), New Zealand (Coutts and Forrest 2007; Smith et al. 2014; Watts et al. 2019), and both coasts of North America (Bullard et al. 2007; Valentine et al. 2007; Cohen et al. 2011; Clancy 2015). Its spread is forecasted to cause significant ecological and economic damage (Coutts & Forrest, 2007; Valentine et al., 2007), as it readily overgrows native eelgrasses (Carman & Grunden 2015; Long

and Grosholz 2015), sessile invertebrate fauna (Valentine 2007a, 2007b; Lengyel et al. 2009), the eggs of commercial fish species such as the Pacific Herring (Cohen et al. 2015), and aquaculture equipment (Switzer et al. 2007; Bullard et al. 2007).

Chimerism in *D. vexillum* can arise through a variety of mechanisms such as allogeneic fusion of mature genetically compatible colonies (Clancy 2015; Smith et al. 2012), larval settlement on colonies, and somatic mutations which are passed on during asexual reproduction (Sommerfeldt et al., 2003) (See Figure 1). Field experiments have found average fusion rates as high as 80% in some introduced populations (Smith et al., 2012), and a positive relationship between colony relatedness and fusion compatibility (Smith et al. 2012; Clancy 2015). In many colonial species, genetically-based self/nonself recognition, or allorecognition, systems permit allogeneic fusion only between related colonies (Grosberg & Quinn, 1988; Lakkis, Dellaporta, & Buss, 2008; Nicotra et al., 2009; Nydam & DeTomaso, 2011; Sabbadin, Zaniolo, & Ballarin, 1992). As recognition specificity derives largely from high levels of polymorphism at allorecognition loci (Grosberg & Quinn, 1988; Nydam & DeTomaso, 2011; Nydam, Taylor, & De Tomaso, 2013; Stoner & Weissman, 1996; Yund & Feldgarden, 1992), genetic bottlenecks can disrupt allorecognition specificity, potentially resulting in populations with a diminished ability to discriminate between related and unrelated colonies, as seen in invasive ants (Tsutsui, Suarez, and Grosberg 2003; Payne, Tillberg, and Suarez 2004), but not necessarily in tunicates (e.g., Wang 2011).

Changes in allorecognition specificity following genetic bottlenecks or other demographic shifts can have far-reaching ecological effects, particularly in the context of invasions (Ben-Shlomo 2017; Payne, Tillberg, and Suarez 2004). This phenomenon is exemplified by the highly invasive Argentine ant, *Linepithema humile*, which forms "super-colonies" throughout its introduced range due to introduction bottlenecks that are thought to have compromised allorecognition specificity in workers (Tsutsui et al. 2001). The emergence of unicoloniality in nonnative populations of *L. humile* and other introduced ants (Tsutsui & Suarez 2003) demonstrates the potential role of allorecognition variation in facilitating biological invasions. Unlike ants, which utilize a combination of behavioral and acquired chemical cues for colony recognition (Brandt et al. 2009; Tsutsui 2013), allorecognition in colonial ascidians is primarily genetically mediated (Cadavid et al., 2004; De Tomaso et al., 2005; Sabbadin et al., 1992; Scofield & Nagashima, 1983), presenting researchers with an opportunity to explore the relationship between chimerism and genetic diversity within and among colonies.

Growing interest in the ecological effects of both inter- and intra-colonial genetic structure in invasive species highlights the need to apply sampling methods that are sensitive enough to detect chimerism even in highly inbred or recently bottlenecked populations. Natural chimerism in *D. vexillum* has previously been measured at frequencies ranging from 12% (Smith 2012) to 48% (Watts, Hopkins, and Goldstien 2019). Previous studies have used samples of colony tissue containing multiple zooids and assessed chimerism based on the presence of more than two alleles at a single

microsatellite locus (Smith 2012; Clancy 2015; Watts et al. 2019). This method provides conservative estimates of chimerism, as it cannot account for molecular masking of chimerism due to homozygosity or spatial grouping of genets within chimeras. A recent study demonstrating that the genotypic composition of chimeric *D. vexillum* colonies may vary spatially and temporally (Weinberg, Clancy, and Cohen, *in press*) further suggests that chimera detection may be subject to influence by a combination of spatial, temporal, and ecological factors.

In this study, we use a novel sampling protocol to measure chimerism and genetic diversity in two invasive *D. vexillum* populations in the Northwest Pacific. Samples from a population in Half Moon Bay (HMB), CA, were collected from Pillar Point Harbor, a small (~47,000 m^2), active marina which experiences a significant amount of small boat traffic from recreational and fishing vessels. In Pillar Point harbor, 73% of transient small boat traffic originates from San Francisco Bay (Zabin et al. 2014), an international shipping hub that may serve as a source for secondary introductions of nonnative species. A second population was sampled from the Umpqua Triangle in Winchester Bay, OR, an area of approximately 200,000 m^2 that is largely enclosed by a concrete jetty and hosts a commercial oyster farm (Figure 2). Boats cannot enter the Triangle, and the jetty impedes water exchange with the surrounding bay, limiting larval dispersal and immigration. *D. vexillum* chimeras have previously been documented in HMB (Clancy 2015), while no previous genetic studies of the Umpqua Triangle population have been performed. In both locations, we examined: i) the frequency of chimerism ii) the extent to which

homozygosity and spatial separation of genets can mask molecular estimates of chimerism, and iii) whether there is higher relatedness among zooids within chimeras than among colonies in the population. Characterizing the prevalence and patterns of chimerism in invasive marine species is an essential first step towards developing a comprehensive understanding of the complex relationship between genetic bottlenecks, allorecognition specificity, chimerism, and the population dynamics of introduced species.

2. Methods

2.1 Sample Collection

Field collections: Sections of *D. vexillum* colonies measuring >5 cm^2 were collected from Pillar Point Harbor in HMB, CA between November 2017 and September 2018. Complete pieces of colony tissue were immediately stored in 95% ethanol until DNA extraction. All colonies were collected from separate dock slips to reduce the likelihood of sampling clones. Colonies representing a range of morphologies including flat mats and rugose mats were collected (see Figure 3). *D. vexillum* samples from the Umpqua Triangle Aquaculture Farm, Oregon, were collected by SCUBA by the Oregon Department of Fish and Wildlife in November 2016. One to three fragments were collected from each colony and preserved in 95% ethanol until DNA extraction.

Zooid isolation and spatial assignment: The genetic composition of *D. vexillum* colonies was measured by removing the thoraces of seven zooids from each colony and individually genotyping each thorax. Only the thoraces were sampled to avoid gonadal germline tissue. All zooid isolation and sampling procedures were performed under a dissecting microscope using flame-sterilized tools and dissecting trays that had been cleaned with bleach and ethanol prior to use.

Immediately prior to sampling, whole sections of ethanol-preserved tissue were submerged in deionized water and cleaned to remove dirt and commensal organisms. Cleaned colonies were then transferred to a new dissecting tray filled with fresh deionized water, where the remaining procedures were performed. Each piece of colony tissue was measured with calipers and the seven locations where zooids would be sampled were marked with flame-sterilized metal pins (see below for procedures used to select sampling location) and photographed.

Continuous pieces of colony tissue: For the purposes of consistent sampling, all colonies were treated as two-dimensional surfaces. Colonies were assigned x and y axes, and oriented so that the x axis corresponded with the maximum length of the colony where possible (see Figure 3). Sampling x-coordinates were then determined based on the maximum length of the colony, so that zooids were sampled at equal intervals along the assigned x-axis. For example:

$$\text{interval}=x_{max}/6$$

$$x_1=0$$

$$x_2=x_1+\text{interval}$$

$$x_3=x_2+\text{interval}...x_7=x_{max}$$

The length of the colony perpendicular to each x-interval was measured and used as the maximum possible y-value for that sample location (y_{max}). The R function runif() (R Core Development Team 2013), was applied to each y_{max} value to randomly select the y-value corresponding to each x-interval. This method was used to generate a series of 7 coordinates for each colony, from (x_1, y_1) through (x_7, y_7), in which x_1 was always equal to 0 and x_7 was always equal to the maximum length of the colony (x_{max}). Figure 3 shows examples of zooids sampled according to this method. When computer-generated sampling coordinates fell on a lobe, tendril or other 3-dimensional feature, the sample was taken from the closest estimated location, and its location as well as other morphological features of the three-dimensional structure were carefully noted.

Colonies sampled in three or more fragments: Due to limitations in the collection and transport of colonies from Umpqua Triangle, OR, the samples collected from single colonies were sometimes in 2-3 separate fragments. A modified version of the protocol described above was applied to select samples from fragmented colonies so that the 7 samples were approximately evenly spaced within the fragment and at least two samples were taken from each physically continuous fragment. The number of zooids sampled per

fragment was determined based on the relative sizes of all fragments in a sample. For example, in a sample split into three fragments, sampling locations were determined as follows:

Fragment 1 x_{max}=1.8 cm	$n_{zooids\ sampled}$ =2
Fragment 2 x_{max}=2.6 cm	$n_{zooids\ sampled}$=2
Fragment 3 x_{max}=4.9 cm	$n_{zooids\ sampled}$=3

No samples that were split into 4 or more fragments were used. For fragments from which only two zooids were sampled, locations were adjusted by 5-10 mm towards the center of the fragment, as the equal x-interval approach described above would have resulted in sampling only the edges of the tissue section, which can lead to the overestimation of genetic heterogeneity due to edge effects (Arnaud-Haond et al. 2007).

2.2. Genotyping methods

2.2.1 DNA extraction

Zooid thoraces were removed from the tunic using flame-sterilized forceps, briefly rinsed in deionized water, and placed immediately into an epi-tube containing 180 µL of SDS-based extraction buffer (T1 lysis buffer, Machery-Nagel). To ensure that contact between the forceps and surrounding tunic tissue did not introduce DNA contaminants, each series of dissections included a negative extraction control in which

we punctured the tunic with forceps but did not remove any zooid tissue. We rinsed the forceps in deionized water and dipped them in an epi-tube containing extraction buffer, which was processed alongside the tissue samples (see DNA extraction methods, below). The tissue samples and negative control were then incubated in the extraction buffer with 25 µL Proteinase K at 56° C for 3-4 hours. The remainder of the extraction was completed using the guanidine-based Machery-Nagel Nucleospin column extraction kit, following the manufacturer's protocol.

COI sequencing of all 20 colonies from Umpqua was performed using DNA that had been extracted prior to the sampling of individual zooids for microsatellite analyses. Sections of tissue measuring ~ 1 cm^2 containing both tunic and multiple zooids were macerated into fine pieces (1 mm^2 or smaller) using flame-sterilized forceps and a sterile razor blade. Macerated tissue was placed in an epi-tube with 180 µL T1 Lysis buffer (Machery-Nagel) and 25 µL Proteinase K then incubated at 56° C for 10-12 hours. The remainder of the extraction was completed using the Nucleospin extraction kit (Machery-Nagel) following manufacturer protocols.

2.2.2 Microsatellite sizing and analyses

DNA from each zooid was PCR-amplified using fluorescently labeled primers for 10 published microsatellite markers (Abbott, Ebert, Tabata, & Therriault, 2011). Each locus was amplified individually, with negative controls for both the DNA extraction and PCR included. Amplification was performed in a 10 µL PCR mix containing dH_2O, 10X PE II

Buffer, 1.5 mM $MgCl_2$, 1μg bovine serum albumin, 200 μM dNTPs, 200 nM each of fluorescently labeled Forward and Reverse Primers, 0.5 units Taq Polymerase (New England Biolabs), and 1-2 μL DNA template. The loci dvex03, dvex05, dvex12, dvex23, dvex26, dvex30, and dvex32 were all amplified using a touchdown thermocycler protocol (Abbott et al. 2011). The loci dvex14, dvex20, and dvex42 were similarly amplified but with adjusted annealing temperatures of 62°C, 57°C, and 52°C, respectively. Successful amplification of microsatellites and the absence of contamination was confirmed by gel electrophoresis. Amplified microsatellite PCR products were multiplexed so that up to six loci could be sized on a single run. Fragment analysis was performed on the ABI 3130 Genetic Analyzer at the Estuary and Ocean Science Center, Romberg Tiburon Campus. All microsatellites were scored by eye using Geneious v. 11.1.5 (BioMatters). Sizing was repeated for 10% of randomly-chosen samples at each locus to ensure consistency. PCR and sizing of informative loci was also repeated for all samples showing chimerism to confirm variant genotypes.

2.2.3 COI sequencing and analysis

A 586 bp segment of COI was PCR-amplified using the published primers *tun_F* and *tun_R* (Stefaniak et al., 2009) in a 25 μL reaction containing dH20, lx ThermoPol reaction buffer (New England Biolabs), 2 mM $MgCl_2$, 1 μg bovine serum albumin, 800 μM dNTPs, 200 nM of each primer, 0.5 units Taq DNA polymerase (New England Biolabs), and 1 μL of DNA template. The thermocycling protocol began with a 3-minute denaturation at 94°C followed by 32 cycles of 10 seconds at 94°C, 30 seconds at 50°C,

50 seconds at 72°C and a final extension at 72°C for 5 minutes. PCR products were cleaned using the ExoSAP-IT® kit (Affymetrix) according to the manufacturer's protocols, then cycle sequenced with the BigDye Terminator v3.1 kit (Applied Biosystems). Sequencing of both the forward and reverse strands was performed on an ABI 3130 Genetic Analyzer at the Estuary and Ocean Science Center, Romberg Tiburon Campus. Consensus sequences of 478-586 bp were obtained by aligning forward and reverse sequences in Geneious v. 11.1.5. For some samples, short sequencer reads prevented haplotype assignment, so internal primers were designed to amplify a segment of approximately 251 base pairs at the end of the COI sequence.

COI sequences from the 20 Umpqua colonies were then aligned to reference sequences of *D. vexillum* haplotypes 1-6 downloaded from GenBank (accession numbers JF738057-JF738060 and JQ663508- JQ663509). Haplotypes were assigned by comparing polymorphic sites between the known haplotypes and our sequences. Haplotype assignments were confirmed by performing a blastn search of Umpqua sequences to confirm that the haplotype with the highest identity match score was consistent with the assignment based on polymorphic sites.

2.3 Genetic Analyses

2.3.1 Analyses of clonal structure and identity

The R package *poppr* v. 2.8.2 (Kamvar, Brooks, & Grünwald, 2015; Kamvar, Tabima, & Grünwald, 2014) was used to calculate the number of multilocus genotypes

(MLGs) in each population with and without rarefaction to the smallest sample size using 999 permutations. Minimum-spanning networks of colonies in each population were also created in *poppr* v.2.8.2. Genotype probabilities (P_{ID}), allelic richness, fixation indices, F_{ST}, inbreeding coefficients (F_{IS}), and linkage disequilibrium were calculated in GenAlEx v. 6.502 (Peakall & Smouse 2006 and 2012). As some differences between MLGs may be due to somatic mutations or genotyping errors, rather than samples actually originating from separate reproductive events (Waits, Luikart, and Taberlet 2001; Van Oosterhout et al. 2004; Arnaud-Haond et al. 2007), the probability of genotypes within a colony originating from distinct sexual reproductive events (P_{sex}) was calculated with and without correction for F_{IS} with the R package *RClone* v. 1.0.2 (Ballilleul and Arnaud-Haond 2016; Bailleul, Stoeckel, and Arnaud-Haond 2016), an R implementation of the software GenClone (Arnaud-Haond and Belkhir 2007). *RClone* was also used to calculate clonal indices for chimeric colonies, including the number of MLGs per colony, clonal richness (R), and Simpson and Hill indices.

Relatedness among colonies within populations and among genets within chimeric colonies was assessed with the R package *related* (Pew, Muir, Wang, & Frasier, 2015), which uses the relatedness calculation methods implemented in the Windows program *COANCESTRY* (Wang, 2011). To select the most appropriate estimator to compare mean relatedness among genets within chimeras to mean relatedness across the population as a whole, genotypes of 100 parent-offspring pairs were simulated in *related* using the allele frequencies of each population. Mean relatedness coefficients were

calculated for the simulated data set using the estimators of Wang (2002), Ritland (1996), Queller & Goodnight (1989), and Lynch & Ritland. Wang's estimator showed the lowest variance across the 100 replicates and was therefore chosen for the analysis.

As the accuracy of a given relatedness estimator depends upon the reference population used for the calculation above (Taylor 2015; Petit, El Mousadik,and Pons 1998), we also examined the effect of including clonal genotypes in the data set and using either one or both populations as the reference for the analysis. We calculated relatedness using the four methods described above in a data set with and without repeated clonal genotypes. Mean within-colony relatedness values obtained from both data sets were within 10% of each other but were lowest when each genotype was included in the data set only once, so this data set was used for relatedness analysis.

The program *STRUCTURE* v. 2.3.3 (Hubisz, Falush, Stephens, & Pritchard, 2009; Pritchard, Stephens, & Donnelly, 2000) was used to examine population structure among the two sites and among colonies within each site. The optimal number of k-clusters for *STRUCTURE* analysis was determined with the program Structure Harvester (Earl & vonHoldt, 2012), which uses the Evanno method (Evanno, Regnaut, & Goudet, 2005) of comparing log-likelihood means.

2.3.2 Compilation of simulated data sets for comparison of chimera detection

To compare the effectiveness of the fine-scale clone-mapping method with other possible sampling methods for detecting chimerism and identifying MLGs, we simulated two types of alternative data sets with the microsatellite data collected for this study. One of the simulated data sets consisted of "compound" genotypes, which combined all of the alleles found in zooids sampled from that colony. Compound genotypes allowed us to examine the potential molecular masking of chimerism due to homozygosity. Chimeras were determined detectable from compound genotypes if they contained at least one locus with more than two alleles. The other simulated data sets were constructed by randomly sampling the genotypes of 2-7 zooids from each colony, with replacement, for 1,000 bootstrap replicates. To do this, we used the "sample_units" function in *RClone*. The simulated data set was used to examine the number of genotypes found per colony across 1,000 bootstrap replicates, and the number of times chimerism (the presence of more than two genotypes) was detected with each number of sample units. The data set of compound genotypes and the results of the simulated samples can be found in the supplementary materials.

3. Results

3.1 Chimerism and Population Genetics

The fine-scale clone-mapping method, employed here for the first time in *D. vexillum*, revealed chimerism in both *D. vexillum* populations studied. In total, 19/35 colonies sampled were chimeric, and of these nine had three or more genotypes. Chimerism prevalence was 33% (5/15) in HMB and 70% (14/20) in Umpqua. All of the colonies with three or more genotypes were from the Umpqua population (Figure 4). A two-sided χ^2 test for proportions indicated significant differences in the proportion of chimeras between the two sites before and after Yate's continuity correction ($p=7.644 \times 10^{-6}$ before correction, $p=3.943 \times 10^{-5}$ after correction). P_{ID} values for HMB and Umpqua were $1.5x 10^{-6}$ and 1.2×10^{-6}, respectively, indicating that the markers used had sufficient power to discriminate among individuals based on MLGs.

Both the frequency of chimerism and the pattern of genetic heterogeneity varied between populations. The list of unique MLGs and the location of each MLG in all of the chimeric colonies identified in this study can be found in supplementary file **S1**. In HMB, a total of 20 unique genotypes were found in the 15 colonies sampled, with an average of 1.33 ± 0.126 MLGs per colony and a maximum of two genotypes in a single colony. In four of the five chimeric colonies from HMB, zooids with genotypes that deviated from the primary genotype were found on the edges of the sampling area. All chimeras from HMB showed asymmetrical distribution in the number of zooids with each genotype, with 5-6 zooids of the primary genotype, and 1-2 zooids with the alternative genotype. In Umpqua, a total of 49 unique MLGs were identified (41.25 ± 2.11 after rarefaction), with

a maximum of six MLGs in a single colony (UMP6), and an average of 2.7 ± 1.625 MLGs per colony across the population. Three MLGs were repeated in different colonies, which may have resulted from the accidental collection of separated fragments of the same clone. In two of the colony pairs with shared MLGs, one of the colonies was chimeric, so the genotype of one colony was repeated in some, but not all, zooids in the second colony.

Genetic patterns across populations: Unbiased expected heterozygosity (uH_E) values for HMB and Umpqua were 0.555 ± 0.032 and 0.565 ± 0.023 and fixation indices (F) were 0.113 ± 0.077 and 0.004 ± 0.054, respectively. After applying Nei and Chesser's corrections for inbreeding and small population size (Nei & Chesser, 1983), G_{IS} values for individual loci ranged from -0.295 to 0.275, with a mean G_{IS} across all loci of 0.045 ± 0.053, although the estimate was not significant after 999 permutations (p=1). G_{ST}, with Nei and Chesser and Nei corrections, ranged from -0.009 to 0.827, with a mean of 0.228 +/= 0.094 across all loci and was significant after 999 permutations (p=0.001).

The Evanno method showed strong support for k=2 clusters in the *STRUCTURE* analysis. Data visualization showed high levels of population structure in between Umpqua and Half Moon Bay (Figure 5). Samples from Umpqua formed a single, well-defined cluster, while HMB appeared to be a mix of both clusters.

COI haplotype diversity in Umpqua: A total of four COI haplotypes were found in Umpqua: haplotypes 2, 3, 5, and 6 (Figure 6). Haplotype 6 was the most common,

appearing in 12 colonies, followed by haplotypes 2 and 3, which appeared in three and four colonies, respectively. Haplotype 5 was found in one colony. *D. vexillum* COI has been found to form three mitochondrial clades (Stefaniak et al., 2012), although only Clade A has been reported outside of the native range of Japan (Smith et al., 2014, 2012; Stefaniak et al., 2009, 2012). All haplotypes found in Umpqua belonged to this clade.

Genetic relationships within and among colonies: In both locations, relatedness was significantly higher within chimeric colonies than in the population as a whole (Figures 7a and 7b) as determined by a one-tailed Wilcoxon rank sum test (Umpqua: $p<<0.001$, HMB: $p<0.05$). While within-population relatedness values varied depending on the reference data set, relatedness within chimeras remained significantly higher than total population relatedness regardless of the reference data set used. When populations were analyzed together (Figure 7a), mean relatedness within chimeras in Umpqua was 0.473 ± 0.056 and 0.241 ± 0.021 among all colonies. In HMB, mean relatedness within chimeras was 0.32 ± 0.166, and -0.046 ± 0.131 in the total population when the full data set was analyzed. When the analysis was performed with the focal population as the reference population, total mean relatedness in Umpqua was -0.018 ± 0.017 and 0.523 ± 0.08 within chimeras, while in Half Moon Bay total relatedness was -0.131 ± 0.047 and within-chimera relatedness was 0.166 ± 0.803.

Minimum spanning networks computed by dissimilarity and by Bruvo distance reflected this pattern. MLGs from the same chimeric colony were frequently linked

together, although exceptions to this were observed in both HMB and Umpqua (See Figures 8a and 8b).

3.2 Comparison of sampling methods for chimera detection

Compound data set: In the data set of combined genotypes, 14 colonies showed chimerism based on the presence of more than two alleles at a single locus. In three colonies from Umpqua and two from HMB, chimerism was masked by homozygosity. Paired T-tests showed significant differences between the number of chimeras detectable from compound genotypes and the true number of chimeras ($p=0.021$) as well as the average number of variable loci identifiable from compound genotypes compared with the true number of variable loci ($p<<0.001$). The true number of variable loci ranged from 1-6 (mean 2.4 ± 0.8763) in the five masked chimeras, and 1-9 (mean 5.429 ± 0.653) in the detectable chimeras. In the detectable chimeras, however, the mean number of variable loci identifiable in the compound genotype was 2.533 ± 0.4508, approximately half of the true number of variable loci.

Data sets of random samples: We examined six different sampling simulations, in which between 2-7 genotypes were randomly sampled, with replacement, from each colony. Each of the six simulated data sets represented results from sampling an increasing

number of units from the true data set. We observed linear relationships between the number of sampling units, the average number of chimeras detected (r^2=0.959-1, p<<0.001), and the average number of MLGs detected (Figures 9a-9b). When 7 simulated units were sampled, all 19 chimeras were identified 100% of the time (Figure 9c). A one-tailed t-test suggested significant differences between the true number of chimeras and the mean number of chimeras identified in simulated data sets of 2-6 samples ($p<2.2 \times 10^{-16}$).

4. Discussion

4.1 Chimerism and Population Genetic Structure

Chimerism was significantly more common in Umpqua than HMB, and the Umpqua population had more genotypes overall and per colony. *D. vexillum* was first reported in the Umpqua Triangle in 2010 (Rumrill, Teem, & Boatner, 2014) and is therefore a recent introduction compared to HMB, where *Didemnum* sp.A, later identified as *D. vexillum* (Stefaniak et al., 2009), was first recorded in 1997 (Bullard et al., 2007). The COI haplotype diversity in Umpqua was similar to that found in other populations of *D. vexillum* on the Pacific Northwest Coast of North America (Stefaniak et al 2012; Clancy 2015). In previous genetic studies of *D. vexillum,* only these same four haplotypes were found in HMB and San Francisco Bay (Clancy 2015), although one additional

haplotype, haplotype 1, has also been reported in the region (Stefaniak et al. 2012; Clancy 2015). The COI locus has been reported to show evidence of chimerism in some invasive ascidian populations (Sheets, Cohen, Ruiz, & da Rocha, 2016), although no apparent COI heterozygosity, suggestive of chimerism, was present in the samples sequenced for this study.

The *STRUCTURE* analysis showed that samples from Umpqua and HMB consistently clustered into two separate groups, with Umpqua consistently forming a single, well-defined cluster. This grouping was somewhat less pronounced in HMB, where several samples were assigned with equal probability to the HMB or Umpqua population. This may indicate that the Umpqua population is in fact isolated, while the HMB population continues to experience admixture with other populations on the Pacific Northwest coast.

Given the positive relationship between relatedness and fusion success in *D. vexillum* previously shown by Clancy (2015) and Smith (2012), we expected that the population with lower genetic diversity would also show a higher incidence of chimerism. This did not appear to be the case, as genetic diversity in Umpqua measured by H_E is not significantly lower than what has been found in other *D. vexillum* populations (Clancy 2015; Smith 2012; Stefaniak et al. 2012) and did not correspond to greater inbreeding, as measured by F. Previous genetic studies of *D. vexillum* in HMB found somewhat higher levels of diversity than we observed here (H_E=0.646 in Clancy

2015 polyploid analysis of 31 colonies; H_E=0.555 in this study). The similarity in H_E values between HMB and Umpqua may have been an artifact of our sampling strategy, as our data set contained about twice as many genotypes from the Umpqua population than HMB (Umpqua=41 after rarefaction; HMB=20) despite a comparable number of colonies sampled from each location (Umpqua=20; HMB=15). This may be due to the high number of genotypes that were found in some colonies from Umpqua. It is also possible that the Umpqua population was separated from an outbred source population too recently for genetic evidence of inbreeding to have accumulated. The reproductive season of *D. vexillum* typically begins when water temperatures increase and lasts for 3-5 months in the northern hemisphere (Fletcher, Forrest, Atalah, & Bell, 2013). As *D. vexillum* larvae take from 3.5-5 months to reach sexual maturity (Fletcher et al., 2013; Valentine, Carman, Dijkstra, & Blackwood, 2009), generation times in Umpqua are expected to be around a year.

Population-level relatedness, however, was significantly different between HMB and Umpqua when calculated from reference data including both populations, with the average relatedness among colonies in Umpqua similar to what might be expected in half-sibling or avuncular relationships. The relatively high level of relatedness among colonies in Umpqua may correlate with higher fusion rates in Umpqua than HMB, which we infer based on the high incidence of chimeras. Although a genetic determinant to fusion in *D. vexillum* has been previously inferred based on the patterns of relatedness in fusing and non-fusing colonies, as of yet no allorecognition loci have been identified.

There is no evidence or expectation that the microsatellite markers used in this study are linked to any such locus, so while genetic diversity at microsatellite loci may not serve as a useful proxy for diversity at allorecognition loci, genotypic diversity could be more informative.

In *Botryllus schlosseri*, which is currently the only colonial ascidian for which an allorecognition locus has been identified and characterized, fusion compatibility is determined based on partial matching at a single genomic region with Mendelian inheritance called the Fusion/Histocompatibility (Fu/Hc) locus (De Tomaso et al., 2005; Sabbadin & Zaniolo, 1979; Scofield & Nagashima, 1983; Taketa et al., 2015; Voskoboynik et al., 2013). Because only one allele must be shared for colonies to fuse, parents and offspring can fuse 100% of the time, and full-siblings can fuse approximately 50% of the time on average. High MLG diversity in Umpqua stemmed from intra-individual diversity, but due to its recent introduction, the Umpqua population may be a rare example of a population where diversity is lower at allorecognition loci than in other parts of the genome. It is also possible that chimeric colonies possess a greater number of allorecognition alleles due to the presence of additional genotypes in the colony and are therefore more likely to share alleles and be able to fuse with adjacent colonies.

4.2 Patterns and Frequency of Chimerism

The frequency of natural chimerism that we saw in Umpqua (70%) is higher than has been reported in other *D. vexillum* populations, which range from 12-48% (Smith

2012; Watts, Hopkins, and Goldstien 2019; Clancy 2015), and this was true even in the comparative data sets of compound genotypes (60% chimeric) and random simulated samples (50%-70% chimeric). To our knowledge, Umpqua has the highest rate of chimerism of any *D. vexillum* population genotyped to date. Because this was the case for all real and simulated data sets from Umpqua, it is likely that our sampling method alone is not responsible for the high prevalence of chimerism here.

Although natural chimerism in Umpqua is more prevalent than what has been reported in other populations, experimental work has shown average fusion rates of up to 80% in introduced populations in New Zealand (Smith et al., 2012), although the frequency of natural chimerism reported in New Zealand *D. vexillum* populations is much lower (Smith, 2012; Watts et al., 2019). This difference may be due to the sensitivity of the molecular detection methods used by these studies, which are not designed to identify different spatial aggregations of clones or chimeras made up of closely related individuals. However, previous research assaying 33 colonies in HMB found a fusion rate of 45% and natural chimerism in 29% of colonies (raw data from Clancy 2015), similar to the 33% prevalence observed in this study. Other environmental and ecological factors such as population density, substrate availability, or disturbance likely contribute to the variation in natural chimerism by influencing the frequency contact between *D. vexillum* colonies. Furthermore, the potential role of environmental factors in determining *D. vexillum* fusion outcomes has yet to be investigated, but variation in the temporal

stability of chimeras can influence estimations of chimerism from genetic samples taken at a single time (Watts et al. 2019, Weinberg et al. *in press*).

Our study does not explicitly address the causes of variation in chimerism across populations of *D. vexillum*, but the relatively high rate of chimerism in Umpqua could be associated with its unique landscape or demographic history. The Triangle, which creates a partial barrier to water exchange with the surrounding bay, may inhibit larval dispersal, potentially promoting larval settlement on histocompatible colonies. The role of larval settlement in maintaining chimerism in this population is particularly intriguing given that mean relatedness within chimeras was 0.473-0.522, and relatedness between parents and offspring or full-siblings is expected to be ~0.5 on average. Preferential settlement of larvae on parental colonies has been reported in *B. schlosseri* (Grosberg and Quinn 1986), and our analyses suggest that frequent larval settlement on parental colonies is a plausible explanation for the high prevalence of chimerism in Umpqua. In addition to high relatedness within chimeras, the lack of spatial aggregation of genets within chimeras in Umpqua also supports the idea that larval settlement contributes to the high rate of chimerism. Unlike allogeneic fusion, in which clonal lineages can remain clustered on either side of the fusion line for several weeks, larval settlement is expected to result in chimeras with a more scattered distribution of genotypes (Figure 1). Zooids of the same genotype were typically clustered together in chimeric colonies from HMB, while no such patterns were evident in chimeras from Umpqua, although fragmentation

of the samples from Umpqua prevented direct comparison of spatial structure between the two locations.

4.3 Comparison of real and simulated data

Comparisons with both the compound and random sample data sets showed that fine-scale sampling can identify more chimeras and a higher number of clonal lineages within chimeras than other approaches. Data from the compound genotypes showed that chimera detection is impeded by molecular masking due to homozygosity or low marker diversity even when the colony area sampled is much larger than typically used. The data set of random samples, which allowed us to simulate the effects of sampling area, also recovered a lower number of chimeras and a lower number of MLGs per colony than were actually present for all simulations of less than seven samples. While the difference in absolute number of chimeras detected was not significant in simulations of five or six samples, the difference in the number of MLGs was. Even for colonies with only two genotypes, the total number of MLGs was recovered 100% of the time only when seven samples were simulated. In a comprehensive sampling effort, the number of MLGs should appear to reach an asymptote at the optimal number of units, and as we did not see this it could be that sampling more zooids would continue to increase the number of MLGs detectable per colony. The number of samples per colony needed to accurately detect MLG diversity likely varies depending on the genetic background of the population and the strength of the molecular markers used.

The aggregations of clonal lineages we observed, particularly in the chimeras from HMB, also point to sampling area and location within the colony as important factors in chimera detection. We suggest that a hybrid approach, with two or more large tissue samples containing multiple zooids collected from different parts of the colony, would be most effective for quantifying the overall frequency of chimerism in a given population. However, this would not account for molecular masking of chimerism, so accurate determination of clonal diversity within a colony would still require the sampling of individual zooids.

Future work will be necessary to further refine approaches to measuring chimerism and intra-colonial genetic structure in ascidians. We have shown here that a sensitive sampling approach developed for the purpose of accurately quantifying the prevalence and patterns of chimerism provides greater resolution than single tissue samples as well as specific information about the number and distribution of MLGs that has not yet been obtained in previous studies of chimerism in *D. vexillum*. Our results contribute to the growing understanding of chimerism in ascidian invasions by showing high levels of variation between populations and identifying factors that influence the accuracy of chimera detection in *D. vexillum*.

Future investigation will be necessary to determine the proximate causes of variation in chimerism rates among populations and the possible effects of this variation on the establishment of non-native *D. vexillum* populations. Quantifying the range of genotypic patterns within and among populations and the habitat parameters associated

with this variation is one step towards understanding the complex interactions underlying variation in allorecognition behavior and invasion success. Studies such as this are integral to developing a complete understanding of how the variation in allorecognition specificity described in terrestrial invasions such as that of the Argentine ant may also contribute to marine invasions. The prevalence of natural chimerism has been found to differ among native and invasive populations of *B. schlosseri* (Ben-Shlomo, Douek, and Rinkevich 2001), and variation in chimerism prevalence in *D. vexillum* populations has been predicted based on differences in fusion frequency between native and invasive populations (Smith et al. 2012; Clancy 2015). Populations with the lowest levels of genetic diversity will be most sensitive to the conservative bias of chimerism estimates that rely on heterozygosity for detection, although it is these same populations that are of the greatest interest to researchers studying the effects of introduction bottlenecks, chimerism rates, and invasion success. Although few other studies have been explicitly designed to measure chimerism with fine-scale sampling, such research efforts will be necessary to provide accurate comparative measurements of chimerism across populations with divergent demographic histories and levels of genetic diversity.

Acknowledgements: We are grateful to all members of the Cohen Lab who helped with this project and provided comments; Steven Rumrill and the Oregon Department of Fish & Wildlife for sample collection from the Umpqua Triangle; Kaitlyn Sapkos, and Natassja Punak for help with sample and data collection from Half Moon Bay; Darragh Clancy for data access and contributions to the development of methods; Benson Chow for help with sample collection from Half Moon Bay and valuable comments; Dr.

Colleen Ingran and Dr. Eric Routman for valuable suggestions, comments, and discussions. **Funding**: Support for this project came from the Estuary & Ocean Science Center (EOS) and San Francisco State University (SFSU) Department of Biology & College of Science and Engineering; Achievement Rewards for College Scientists (ARCS) to RW in 2017-2018 and 2018-2019; SFSU IRA support to RW in 2017 and 2018; CSU Council on Ocean Affairs, Science, and Technology (COAST) Graduate Research Award to RW 2017; EOS Center gene lab is funded by NSF FML 0435033, EOS Center, SFSU College of Science and Engineering, and other anonymous donors.

References:

Abbott, C. L., Ebert, D., Tabata, A., & Therriault, T. W. (2011). Twelve microsatellite markers in the invasive tunicate, Didemnum vexillum, isolated from low genome coverage 454 pyrosequencing reads. *Conservation Genetics Resources*, *3*, 79–81. https://doi.org/10.1007/s12686-010-9294-2

Arnaud-Haond, S., Duarte, C. M., Alberto, F., & Serrão, E. A. (2007). Standardizing methods to address clonality in population studies. *Molecular Ecology*, *16*(24), 5115–5139. https://doi.org/10.1111/j.1365-294X.2007.03535.x

Arnaud-Haond, Sophie, & Belkhir, K. (2007). GENCLONE: A computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Molecular Ecology Notes*, *7*(1), 15–17. https://doi.org/10.1111/j.1471-8286.2006.01522.x

Bailleul, D., Stoeckel, S., & Arnaud-Haond, S. (2016). RClone: a package to identify MultiLocus Clonal Lineages and handle clonal data sets in r. *Methods in Ecology and Evolution*, *7*(8), 966–970. https://doi.org/10.1111/2041-210X.12550

Bancroft, F. W. (1903). Variation and Fusion of Colonies in Compound Ascidians. *Proceedings of the California Academy of Sciences*, *3*(5).

Becheler, R., Cassone, A. L., Noël, P., Mouchel, O., Morrison, C. L., & Arnaud-Haond, S. (2017). Low incidence of clonality in cold water corals revealed through the novel use of a standardized protocol adapted to deep sea sampling. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *145*(December 2015), 120–130. https://doi.org/10.1016/j.dsr2.2015.11.013

Ben-Shlomo, R., Douek, J., & Rinkevich, B. (2001). Heterozygote deficiency and chimerism in remote populations of a colonial ascidian from New Zealand. *Marine*

Ecology Progress Series, *209*(June 2014), 109–117. https://doi.org/10.3354/meps209109

Ben-Shlomo, Rachel. (2017). *Invasiveness, chimerism and genetic diversity*. *38*(1), 42–49. https://doi.org/10.1111/ijlh.12426

Blanquer, A., & Uriz, M. J. (2011). Living together apart: The hidden genetic diversity of sponge populations. *Molecular Biology and Evolution*, *28*(9), 2435–2438. https://doi.org/10.1093/molbev/msr096

Bullard, S. G., Lambert, G., Carman, M. R., Byrnes, J., Whitlatch, R. B., Ruiz, G., … Heinonen, K. (2007). The colonial ascidian *Didemnum* sp. A: Current distribution, basic biology and potential threat to marine communities of the northeast and west coasts of North America. *Journal of Experimental Marine Biology and Ecology*, *342*(1 SPEC. ISS.), 99–108. https://doi.org/10.1016/j.jembe.2006.10.020

Cadavid, L. F., Powell, A. E., Nicotra, M. L., Moreno, M., & Buss, L. W. (2004). An invertebrate histocompatibility complex. *Genetics*, *167*(1), 357–365. https://doi.org/10.1534/genetics.167.1.357

Castillo, D. I., Queller, D. C., & Strassmann, J. E. (2011). Cell condition, competition, and chimerism in the social amoeba Dictyostelium discoideum. *Ethology Ecology and Evolution*, *23*(3), 262–273. https://doi.org/10.1080/03949370.2011.568526

Chang, E. S., Orive, M. E., & Cartwright, P. (2018). Nonclonal coloniality: Genetically chimeric colonies through fusion of sexually produced polyps in the hydrozoan Ectopleura larynx . *Evolution Letters*, *2*(4), 442–455. https://doi.org/10.1002/evl3.68

Clancy, D.L. (2015). Examining Genetic Diversity and Fusion Abilities of an Invasive Colonial Ascidian. Master's Thesis. San Francisco State.

Cohen, C. S., McCann, L., Davis, T., Shaw, L., & Ruiz, G. (2011). Discovery and significance of the colonial tunicate Didemnum vexillum in Alaska. *Aquatic Invasions*, *6*(3), 263–271. https://doi.org/10.3391/ai.2011.6.3.03

Coutts, A. D. M., & Forrest, B. M. (2007). Development and application of tools for incursion response: Lessons learned from the management of the fouling pest Didemnum vexillum. *Journal of Experimental Marine Biology and Ecology*, *342*(1 SPEC. ISS.), 154–162. https://doi.org/10.1016/j.jembe.2006.10.042

Crawford, K. M., & Whitney, K. D. (2010). Population genetic diversity influences colonization success. *Molecular Ecology*, *19*(6), 1253–1263. https://doi.org/10.1111/j.1365-294X.2010.04550.x

De Tomaso, A. W., Nyholm, S. V., Palmeri, K. J., Ishizuka, K. J., Ludington, W. B., Mitchel, K., & Weissman, I. L. (2005). Isolation and characterization of a protochordate histocompatibility locus. *Nature*, *438*(7067), 454–459. https://doi.org/10.1038/nature04150

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fletcher, L. M., Forrest, B. M., Atalah, J., & Bell, J. J. (2013). Reproductive seasonality of the invasive ascidian *Didemnum vexillum* in New Zealand and implications for shellfish aquaculture. *Aquaculture Environment Interactions*, *3*(3), 197–211. https://doi.org/10.3354/aei00063

Foster, K. R., Fortunato, A., Strassmann, J. E., & Queller, D. C. (2002). The costs and benefits of being a chimera. *Proceedings of the Royal Society B: Biological Sciences*, *269*(1507), 2357–2362. https://doi.org/10.1098/rspb.2002.2163

Frank, M. H., & Chitwood, D. H. (2016). Plant chimeras: The good, the bad, and the 'Bizzaria.' *Developmental Biology*, *419*(1), 41–53. https://doi.org/10.1016/j.ydbio.2016.07.003

Gauthier, M., & Degnan, B. M. (2008). Partitioning of genetically distinct cell populations in chimeric juveniles of the sponge Amphimedon queenslandica. *Developmental and Comparative Immunology*, *32*(11), 1270–1280. https://doi.org/10.1016/j.dci.2008.04.002

Gill, D. E., Chao, L., Perkins, S. L., & Wolf, J. B. (2003). Genetic Mosaicism in Plants and Clonal Animals. *Annual Review of Ecology and Systematics*, *26*(1), 423–444. https://doi.org/10.1146/annurev.es.26.110195.002231

Gröner, F., Lenz, M., Wahl, M., & Jenkins, S. R. (2011). Stress resistance in two colonial ascidians from the Irish Sea: The recent invader Didemnum vexillum is more tolerant to low salinity than the cosmopolitan Diplosoma listerianum. *Journal of Experimental Marine Biology and Ecology*, *409*(1–2), 48–52. https://doi.org/10.1016/j.jembe.2011.08.002

Grosberg, R. K., & Quinn, J. F. (1988). The Evolution of Allorecognition Specificity. In *Invertebrate Historecognition* (pp. 157–167). https://doi.org/10.1007/978-1-4613-

1053-2_12

Hart, M. W., & Grosberg, R. K. (1999). Kin Interactions in a Colonial Hydrozoan (Hydractinia symbiolongicarpus): Population Structure on a Mobile Landscape. *Society for the Study of Evolution*, *53*(3), 793–805.

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, *9*(5), 1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Ilan, M., & Loya, Y. (1990). Ontogenetic variation in sponge histocompatibility responses. *Biological Bulletin*, *179*(3), 279–286. https://doi.org/10.2307/1542319

Kamvar, Z., Brooks, J., & Grünwald, N. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Fronteirs in Genetics*, *6*(208). https://doi.org/doi: 10.3389/fgene.2015.00208

Kamvar, Z., Tabima, J., & Grünwald, N. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2:e821*. https://doi.org/2:e281. doi: 10.7717/peerj.281

Lakkis, F. G., Dellaporta, S. L., & Buss, L. W. (2008). Allorecognition and chimerism in an invertebrate model organism. *Organogenesis*, *58*(9), 1–6. https://doi.org/10.1641/b580912

Lambert, G. (2009). Adventures of a sea squirt sleuth: unraveling the identity of Didemnum vexillum, a global ascidian invader. *Aquatic Invasions*, *4*(1), 5–28. https://doi.org/10.3391/ai

Maier, E., Buckenmaier, A., Tollrian, R., & Nürnberger, B. (2012). Intracolonial genetic variation in the scleractinian coral Seriatopora hystrix. *Coral Reefs*, *31*(2), 505–517. https://doi.org/10.1007/s00338-011-0857-9

Monro, K., & Poore, A. G. B. (2004). Selection in Modular Organisms: Is Intraclonal Variation in Macroalgae Evolutionarily Important? *The American Naturalist*, *163*(4), 564–578. https://doi.org/10.1086/382551

Nawrocki, A. M., & Cartwright, P. (2012). A novel mode of colony formation in a hydrozoan through fusion of sexually generated individuals. *Current Biology*, *22*(9), 825–829. https://doi.org/10.1016/j.cub.2012.03.026

Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, *47*, 253–259.

Nicotra, M. L., Powell, A. E., Rosengarten, R. D., Moreno, M., Grimwood, J., Lakkis, F. G., … Buss, L. W. (2009). A Hypervariable Invertebrate Allodeterminant. *Current Biology*, *19*(7), 583–589. https://doi.org/10.1016/j.cub.2009.02.040

Nydam, M. L., & DeTomaso, A. W. (2011). Creation and maintenance of variation in allorecognition loci: Molecular analysis in various model systems. *Frontiers in Immunology*, *2*(DEC), 1–6. https://doi.org/10.3389/fimmu.2011.00079

Nydam, M. L., Taylor, A. A., & De Tomaso, A. W. (2013). Evidence for selection on a chordate histocompatibility locus. *Evolution*, *67*(2), 487–500. https://doi.org/10.1111/j.1558-5646.2012.01787.x

Oka, H., & Watanabe, H. (1957). Colony-specificity in compound ascidians as tested by fusion experiments (a preliminary report). *Proceedings of the Japan Academy*, *33*(10), 657–659.

Ordóñez, V., Pascual, M., Fernández-Tejedor, M., Pineda, M. C., Tagliapietra, D., & Turon, X. (2015). Ongoing expansion of the worldwide invader Didemnum vexillum (Ascidiacea) in the Mediterranean Sea: high plasticity of its biological cycle promotes establishment in warm waters. *Biological Invasions*, *17*(7), 2075–2085. https://doi.org/10.1007/s10530-015-0861-z

Osman, R. W., & Whitlatch, R. B. (2007). Variation in the ability of Didemnum sp. to invade established communities. *Journal of Experimental Marine Biology and Ecology*, *342*(1 SPEC. ISS.), 40–53. https://doi.org/10.1016/j.jembe.2006.10.013

Petit,Remy J, El Mousadik,Abdelhamid, & Pons,Odile. (1998). Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, *12*(4), 844–855. https://doi.org/10.1046/j.1523-1739.1998.96489.x

Pew, J., Muir, P. H., Wang, J., & Frasier, T. R. (2015). related: An R package for analysing pairwise relatedness from codominant molecular markers. *Molecular Ecology Resources*, *15*(3), 557–561. https://doi.org/10.1111/1755-0998.12323

Pineda-Krch, M., & Lehtilä, K. (2004). Costs and benefits of genetic heterogeneity within organisms. *Journal of Evolutionary Biology*, *17*(6), 1167–1177. https://doi.org/10.1111/j.1420-9101.2004.00808.x

Pineda-Krch, Mario, & Poore, A. G. B. (2004). Spatial interactions within modular organisms: Genetic heterogeneity and organism fitness. *Theoretical Population Biology*, *66*(1), 25–36. https://doi.org/10.1016/j.tpb.2004.03.002

Poore, A. G. B., & Fagerström, T. (2000). Intraclonal variation in macroalgae: causes and evolutionary consequences. *Selection*.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure using Multilocus Genotype Data. *Genetics*, *155*, 945–959.

Puill-Stephan, E., van Oppen, M. J. H., Pichavant-Rafini, K., & Willis, B. L. (2012). High potential for formation and persistence of chimeras following aggregated larval settlement in the broadcast spawning coral, Acropora millepora. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1729), 699–708. https://doi.org/10.1098/rspb.2011.1035

Queller, D. C., & Goodnight, K. F. (1989). Estimating Relatedness Using Genetic Markers. *Evolution*, *43*(2), 258–275. https://doi.org/10.2307/2409206

Rinkevich, B. (2004). Environmental split between germ cell parasitism and somatic cell synergism in chimeras of a colonial urochordate. *Journal of Experimental Biology*, *207*(20), 3531–3536. https://doi.org/10.1242/jeb.01184

Ritland, K. (1996). Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research*, *67*(2), 175–185. https://doi.org/10.1017/s0016672300033620

Rosengarten, R. D., & Nicotra, M. L. (2011). Model systems of invertebrate allorecognition. *Current Biology*, *21*(2), R82–R92. https://doi.org/10.1016/j.cub.2010.11.061

Rumrill, S. S., Teem, J. A., & Boatner, R. (2014). *Risk Assessment for the Compound Tunicate Didemnum vexillum in Oregon ' s Marine Waters June 2014 Oregon Invasive Species Council*. (June), 1–33.

Sabbadin, A., & Zaniolo, G. (1979). Sexual differentiation and germ cell transfer in the colonial ascidian Botryllus schlosseri. *Journal of Experimental Zoology*, *207*(2), 289–304. https://doi.org/10.1002/jez.1402070212

Sabbadin, A., Zaniolo, G., & Ballarin, L. (1992). Genetic and cytological aspects of histocompatibility in ascidians. *Bolletino Di Zoologia*, *59*(2), 167–173. https://doi.org/10.1080/11250009209386665

Saito, Y., Hirose, E., & Watanabe, H. (1994). Allorecognition in compound ascidians. *International Journal of Developmental Biology*, *38*(2), 237–247.

Santelices, B., González, A. V., Beltrán, J., & Flores, V. (2017). Coalescing red algae exhibit noninvasive, reversible chimerism. *Journal of Phycology*, *53*(1), 59–69. https://doi.org/10.1111/jpy.12476

Schweinsberg, M., Weiss, L. C., Striewski, S., Tollrian, R., & Lampert, K. P. (2015).

More than one genotype: how common is intracolonial genetic variability in scleractinian corals? *Molecular Ecology*, *24*(11), 2673–2685. https://doi.org/10.1111/mec.13200

Scofield, V. L., & Nagashima, L. S. (1983). Morphology and Genetics of Rejection Reactions between Oozoids from the Tunicate Botryllus schlosseri. *Biological Bulletin*, (165), 733–744.

Sheets, E. A., Cohen, C. S., Ruiz, G. M., & da Rocha, R. M. (2016). Investigating the widespread introduction of a tropical marine fouling species. *Ecology and Evolution*, *6*(8), 2453–2471. https://doi.org/10.1002/ece3.2065

Shenk, M. A. (1991). Allorecognition in the colonial marine hydroid Hydractinia (cnidaria/hydrozoa). *Integrative and Comparative Biology*, *31*(3), 549–557. https://doi.org/10.1093/icb/31.3.549

Smith, K. F. (2012). *Use of genetic methods for determining patterns and processes during marine biological invasions*. Ph.D. Dissertation. Univeristy of Waikato.

Smith, K. F., Abbott, C. L., Saito, Y., & Fidler, A. E. (2014). Comparison of whole mitochondrial genome sequences from two clades of the invasive ascidian, Didemnum vexillum. *Marine Genomics*, *19*, 75–83. https://doi.org/10.1016/j.margen.2014.11.007

Smith, K. F., Stefaniak, L., Saito, Y., Gemmill, C. E. C., Cary, S. C., & Fidler, A. E. (2012). Increased inter-colony fusion rates are associated with reduced COI haplotype diversity in an invasive colonial ascidian Didemnum vexillum. *PLoS ONE*, *7*(1). https://doi.org/10.1371/journal.pone.0030473

Sommerfeldt, A. D., Bishop, J. D. D., & Wood, C. A. (2003). Intraclonal genetic variation : ecological and evolutionary aspects . Chimerism following fusion in a clonal ascidian (Urochordata). *Biological Journal of the Linnean Society*, *79*(November), 183–192.

Stefaniak, L., Lambert, G., Gittenberger, A., Zhang, H., Lin, S., & Whitlatch, R. B. (2009). Genetic conspecificity of the worldwide populations of Didemnum vexillum Kott, 2002. *Aquatic Invasions*, *4*(1), 29–44. https://doi.org/10.3391/ai.2009.4.1.3

Stefaniak, L., Zhang, H., Gittenberger, A., Smith, K., Holsinger, K., Lin, S., & Whitlatch, R. B. (2012). Determining the native region of the putatively invasive ascidian Didemnum vexillum Kott, 2002. *Journal of Experimental Marine Biology and Ecology*, *422–423*, 64–71. https://doi.org/10.1016/j.jembe.2012.04.012

Stoner, D. S., & Weissman, I. L. (1996). Somatic and germ cell parasitism in a colonial

ascidian: possible role for a highly polymorphic allorecognition system. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(26), 15254–15259.

Taketa, D. A., Nydam, M. L., Langenbacher, A. D., Rodriguez, D., Sanders, E., & De Tomaso, A. W. (2015). Molecular evolution and in vitro characterization of Botryllus histocompatibility factor. *Immunogenetics*, *67*(10), 605–623. https://doi.org/10.1007/s00251-015-0870-1

Taylor, H. R. (2015). The use and abuse of genetic marker-based estimates of relatedness and inbreeding. *Ecology and Evolution*, *5*(15), 3140–3150. https://doi.org/10.1002/ece3.1541

Valentine, P. C., Carman, M. R., Dijkstra, J., & Blackwood, D. S. (2009). Larval recruitment of the invasive colonial ascidian Didemnum vexillum, seasonal water temperatures in New England coastal and offshore waters, and implications for spread of the species. *Aquatic Invasions*, *4*(1), 153–168. https://doi.org/10.3391/ai.2009.4.1.16

Valentine, P. C., Collie, J. S., Reid, R. N., Asch, R. G., Guida, V. G., & Blackwood, D. S. (2007). The occurrence of the colonial ascidian Didemnum sp. on Georges Bank gravel habitat - Ecological observations and potential effects on groundfish and scallop fisheries. *Journal of Experimental Marine Biology and Ecology*, *342*(1 SPEC. ISS.), 179–181. https://doi.org/10.1016/j.jembe.2006.10.038

Voskoboynik, A., Newman, A. M., Corey, D. M., Sahoo, D., Pushkarev, D., Neff, N. F., … Weissman, I. L. (2013). Identification of a Colonial Chordate Histocompatibility Complex. *Science*, *341*(6144), 384–387. https://doi.org/10.1126/science.1238036

Wang, J. (2002). An estimator for pairwise relatedness using molecular markers. *Genetics*, *160*(3), 1203–1215.

Wang, J. (2011). Coancestry: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, *11*(1), 141–145. https://doi.org/10.1111/j.1755-0998.2010.02885.x

Watts, A. M., Hopkins, G. A., & Goldstien, S. J. (2019). Chimerism and population dieback alter genetic inference related to invasion pathways and connectivity of biofouling populations on artificial substrata. *Ecology and Evolution*, (September 2018), 3089–3104. https://doi.org/10.1002/ece3.4817

Westerman, E. L., Dijkstra, J. A., & Harris, L. G. (2009). High natural fusion rates in a botryllid ascidian. *Marine Biology*, *156*(12), 2613–2619. https://doi.org/10.1007/s00227-009-1287-x

Yund, P. O., & Feldgarden, M. (1992). Rapid proliferation of historecognition alleles in populations of a colonial ascidian. *Journal of Experimental Zoology*, *263*(4), 442–452. https://doi.org/10.1002/jez.1402630412

Table 1a. Summary of chimerism in Half Moon Bay and Umpqua. N_{cols}, number of colonies sampled; MLGs, the total number of MLGs found in the population; eMLGs, number of multilocus genotypes, rarefied to n=104 samples; C, number of chimeric colonies detected; G/C_{mean}, mean number of genotypes per colony; G/C_{max}, maximum number of genotypes detected in a single colony. All values are shown with ± standard error, where applicable.

Pop	N_{cols}	MLGs	eMLGs	C	G/C_{mean}	G/C_{max}
HMB	15	20	20	5(33%)	1.33 ± 0.126	2
Umpqua	20	49	41.25 ± 2.11	14 (70%)	2.7 ± 1.626	6

Table 1b. Summary of diversity measures for each population. Ar, allelic richness (total over all loci); uHe, Nei's unbiased expected heterozygosity; F_{IS}, fixation index; I, Shannon-Weiner MLG diversity index (Pielou 196); R, clonal diversity index (Dorken & Eckert 2001; Ellstrand & Roose 1987); H, Hill index of clonal diversity (Hurlbert 1971 ; Hill 1973). All values are shown with ± standard error where applicable.

Pop	A_r	uHe	F_{IS}	I	R	H
HMB	45	0.555 ± 0.32	0.113 ± 0.077	2.870	0.184	19.763
Umpqua	38	0.567 ± 0.024	0.004 ± 0.075	3.373	0.345	21.291

Table 2. Summary statistics for HMB and Umpqua. Bolded values are significant at $p<0.01$ after 999 permutations. All values are shown with ± standard error where applicable

Jost D	G''_{ST}	G_{IS}
0.141 ± 0.064	**0.226 ± 0.093**	0.045 ± 0.053

Figure 1. Possible routes to chimerism and expected patterns of genotypic distributions resulting from each.

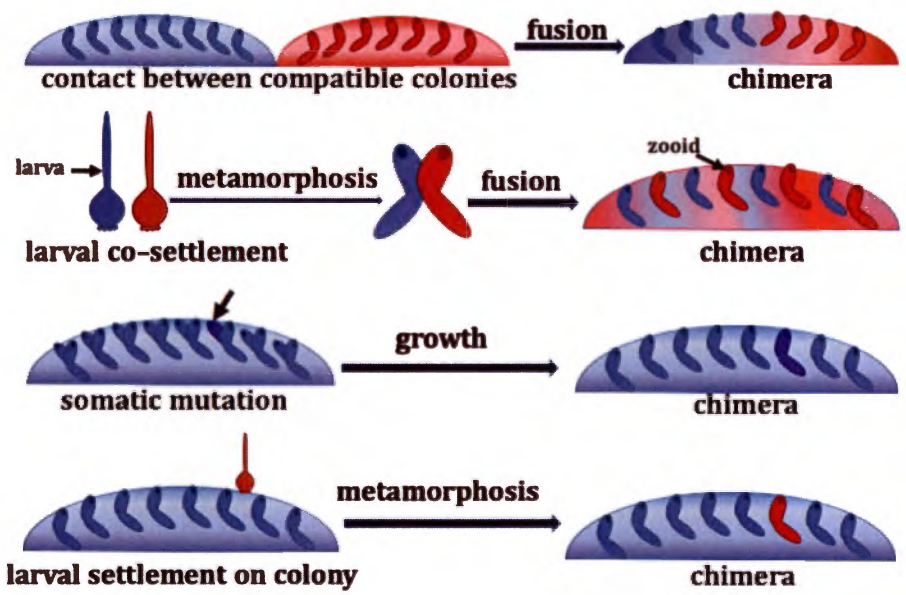
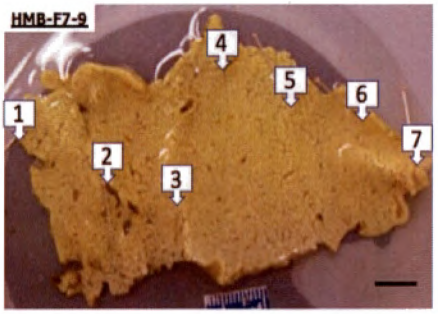


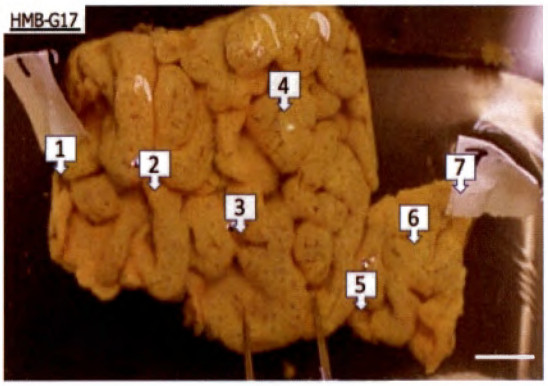
Figure 2. Satellite images and pictures of the Umpqua Triangle, Oregon. The red outline in (A) shows the location of the triangle. Red arrows in (B) and (C) point to the oyster racks where the samples were collected.

Figure 3a-3c. Examples of sampling locations and morphology of selected colonies examined in this study. Numbered white arrows indicate the locations of zooids sampled. Images for all colonies can be found in the **supplementary materials.** A) "Mat" colony from Half Moon Bay. B) "Rugose" colony from Half Moon Bay C) Colony collected in 3 fragments from Umpqua.
All scale bars are 1 cm.

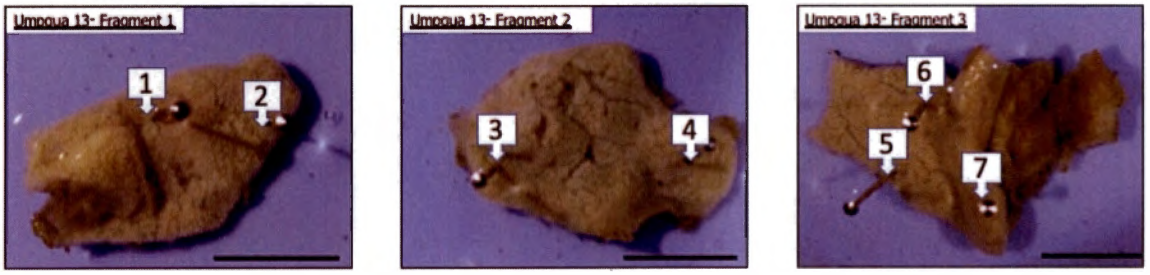
A.



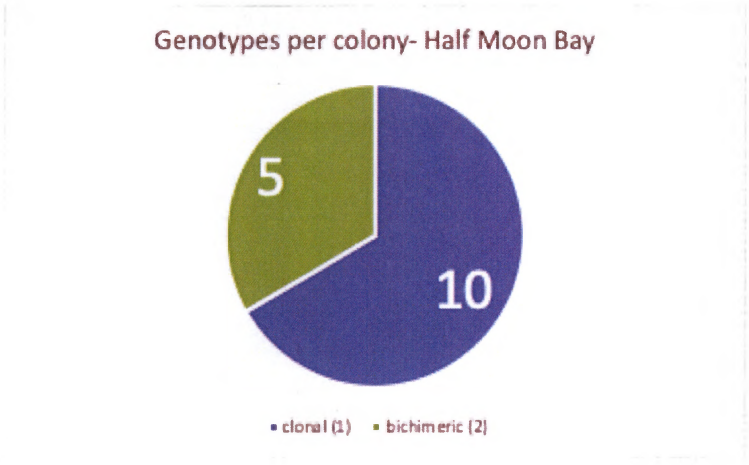
B.



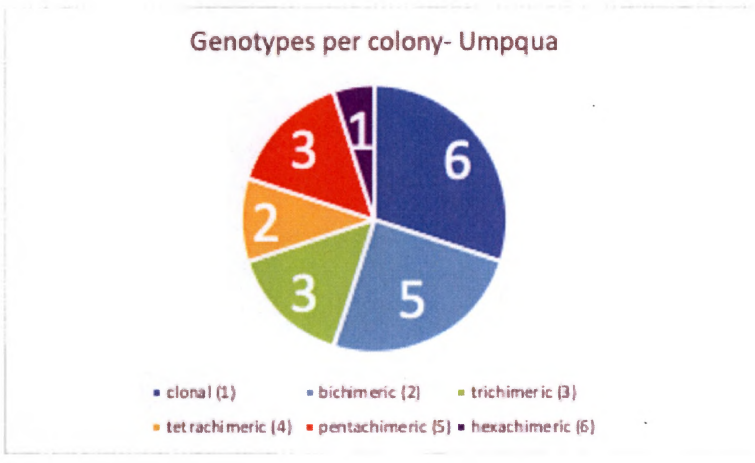
C.



Figures 4a and 4b. Pie charts showing classification of colonies based on number of genotypes for HMB and Umpqua. The legend shows the names of each classification, with the number of genotypes in parentheses. Numbers within the chart are the number of colonies in the category.



A.



B.

Figure 5. Structure plot of Half Moon Bay (1) and Umpqua (2) with K=2 clusters.

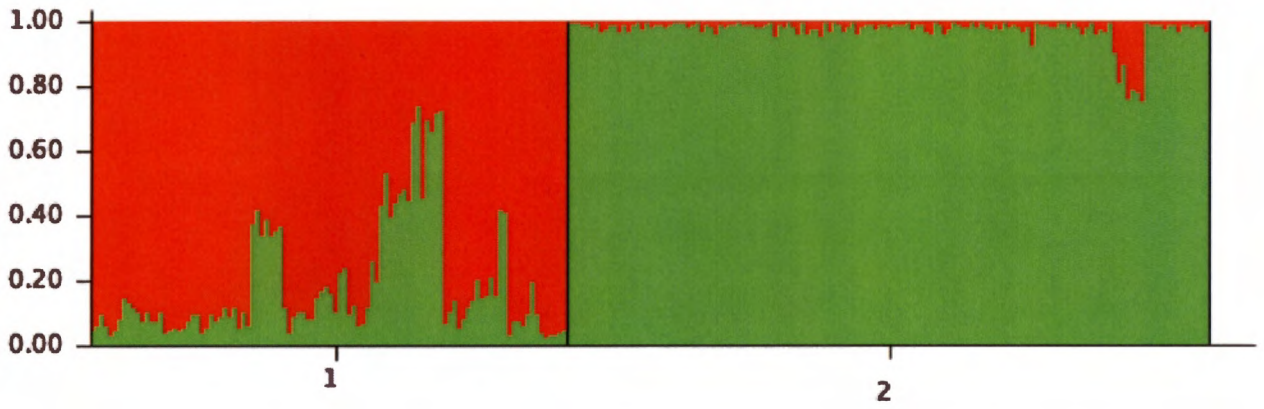


Figure 6. COI Haplotype diversity in Umpqua.

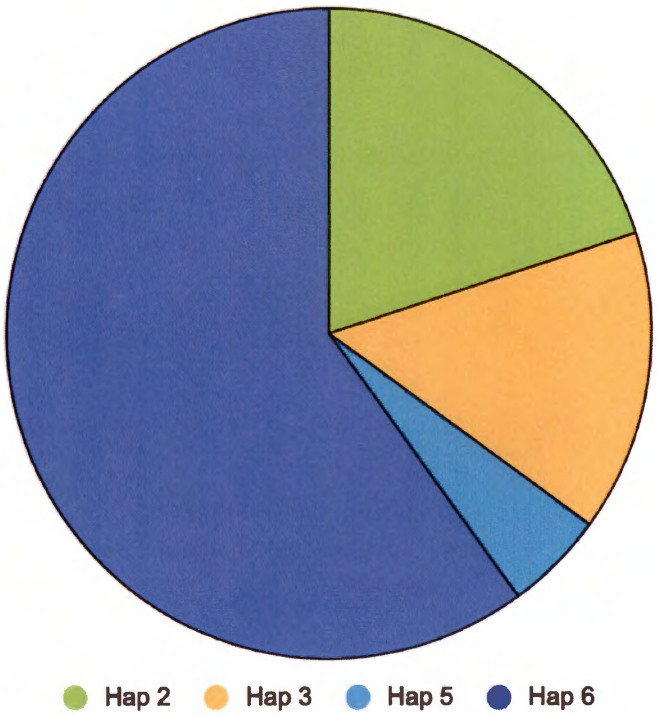
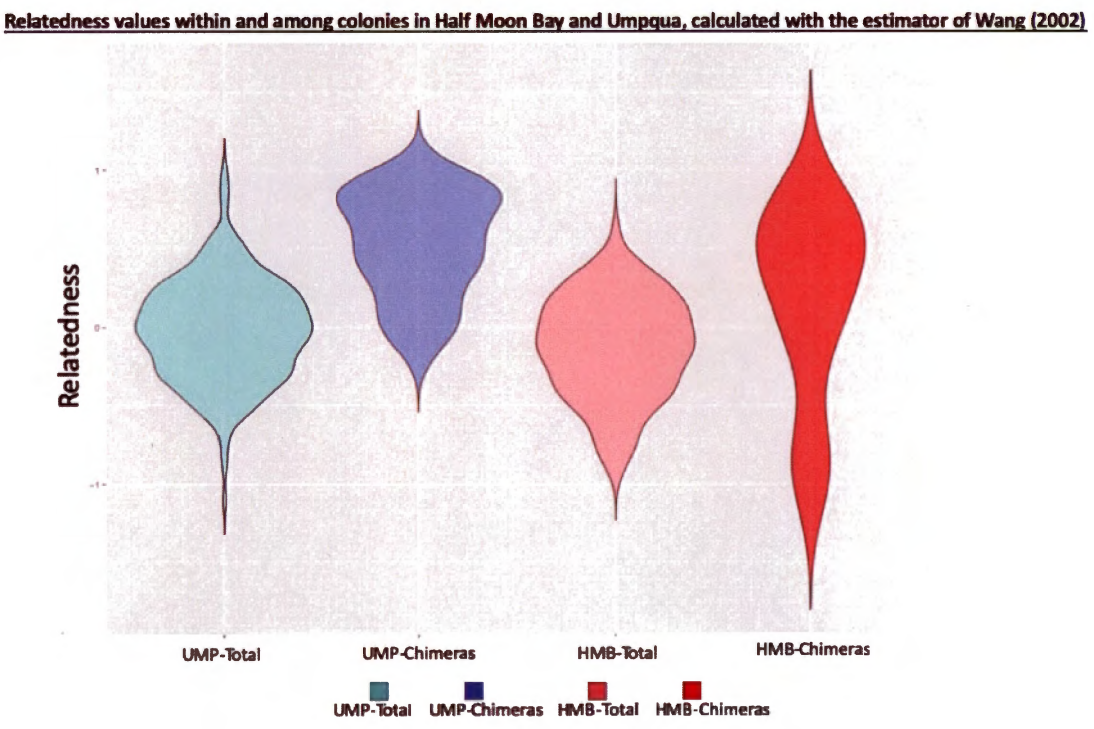
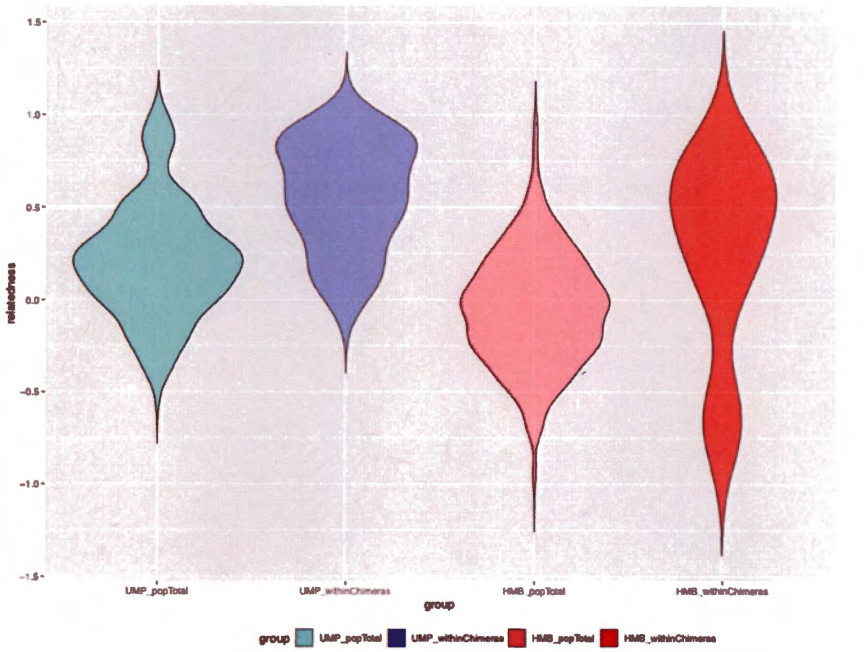


Figure 7: Violin plot showing distribution of mean relatedness values within and among colonies in Umpqua and Half Moon Bay. Relatedness values shown are calculated with the estimator of Wang (2002) and with the focal population used as the reference population for calculations (A) or both populations used as reference (B).

A.

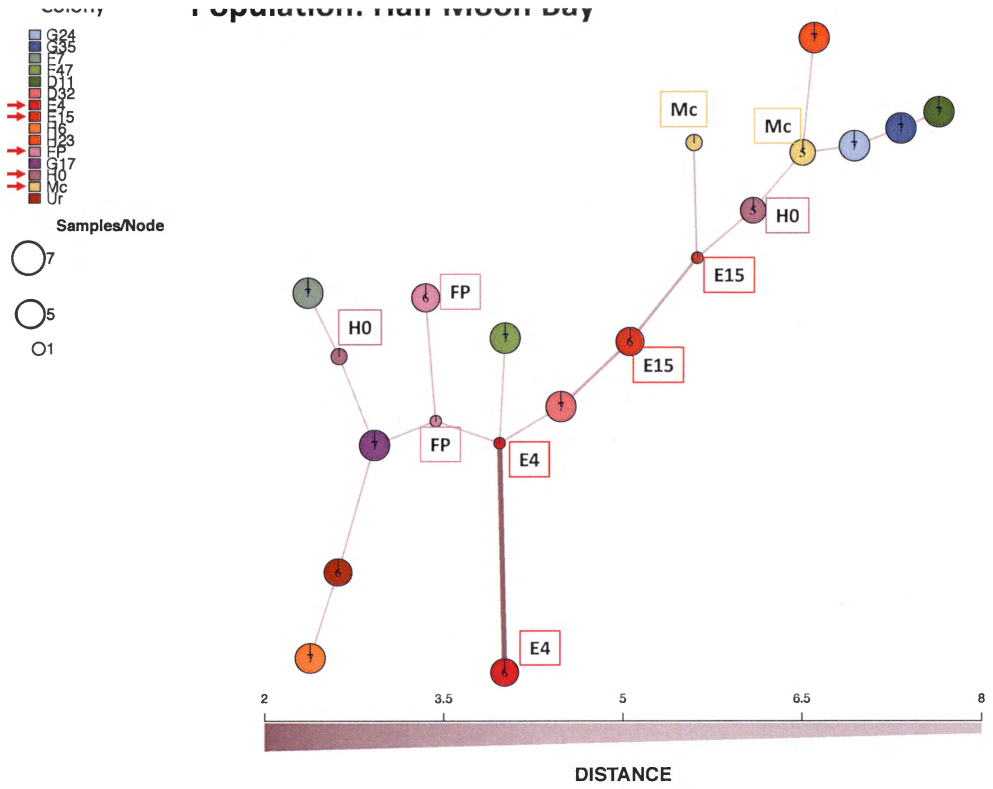


B.


1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
relatedness
UMP_popTotal
UMP_withinChimeras
HMB_popTotal
HMB_withinChimeras
group
group
UMP_popTotal
UMP_withinChimeras
HMB_popTotal
HMB_withinChimeras

Figures 8a and 8b: Minimum spanning networks of dissimilarity distances between MLGs in Half Moon Bay (A) and Umpqua (B) created in *poppr*. Colors correspond to the colonies that samples came from and circles represent unique MLGs. Node width and color is proportional to dissimilarity distance between MLGs. Arrows in the figure legend indicate chimeric colonies, which are also labeled on the network.

A.



B.

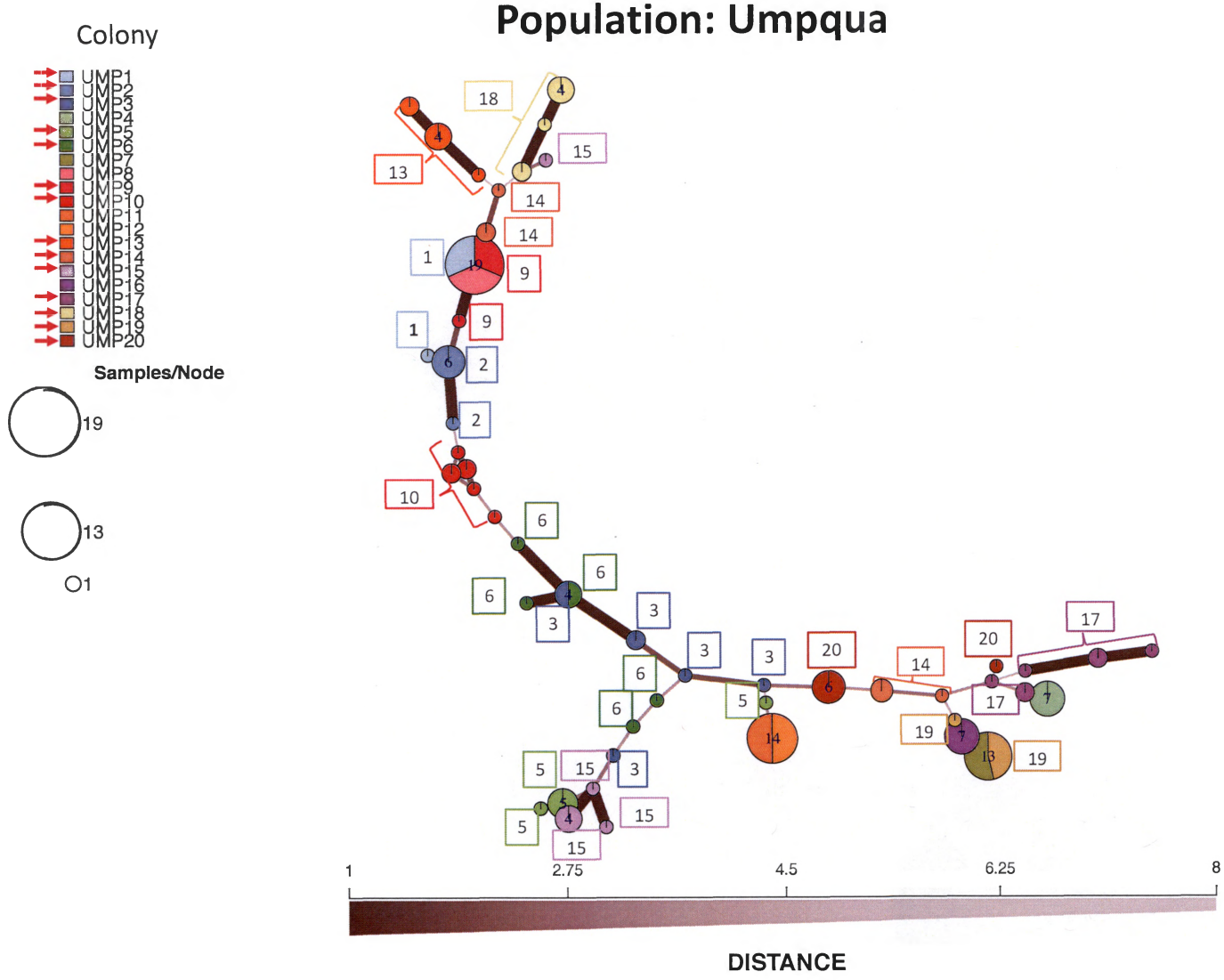

Population: Umpqua
Colony
UMP1
UMP2
UMP3
UMP4
UMP5
UMP6
UMP7
UMP8
UMP9
UMP10
UMP11
UMP12
UMP13
UMP14
UMP15
UMP16
UMP17
UMP18
UMP19
UMP20
Samples/Node
19
13
1
1
2.75
4.5
6.25
8
DISTANCE

Figure 9a. The absolute number of genotypes recovered from each of the simulated sample trials, with points color-coded based on the true number of MLGs in the colony.

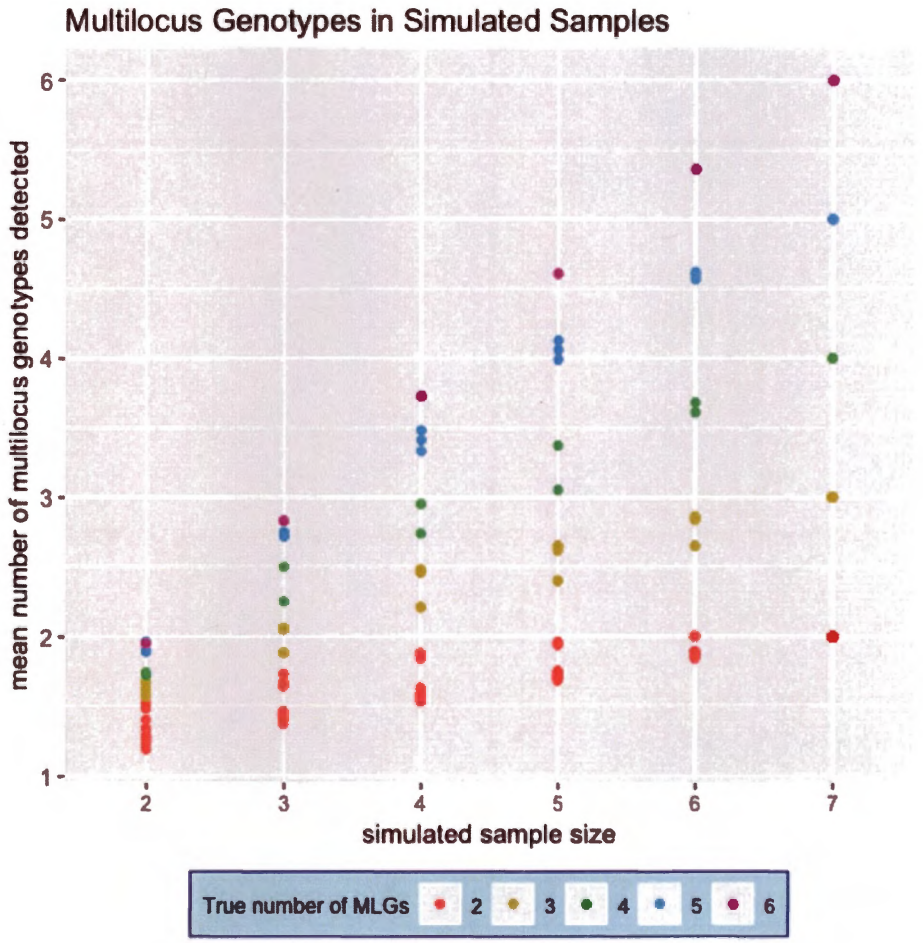


Figure 9b. The proportion of genotypes recovered in each of the simulated samples, separated based on the true number of MLGs in the chimeric colony.

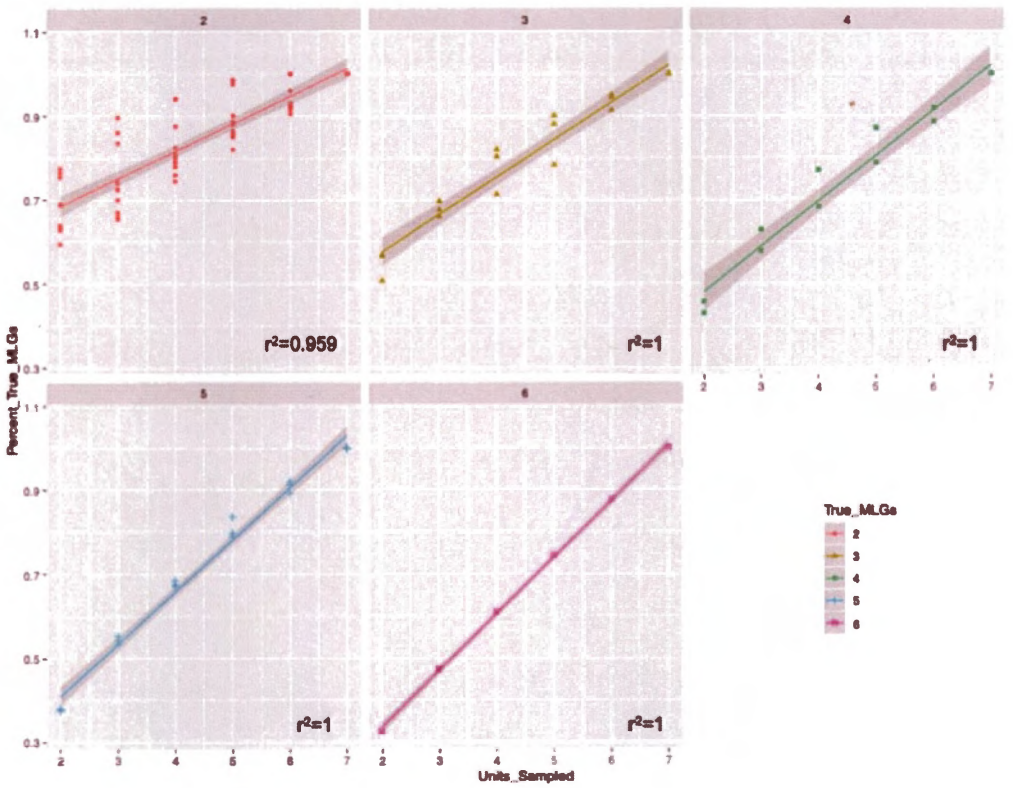
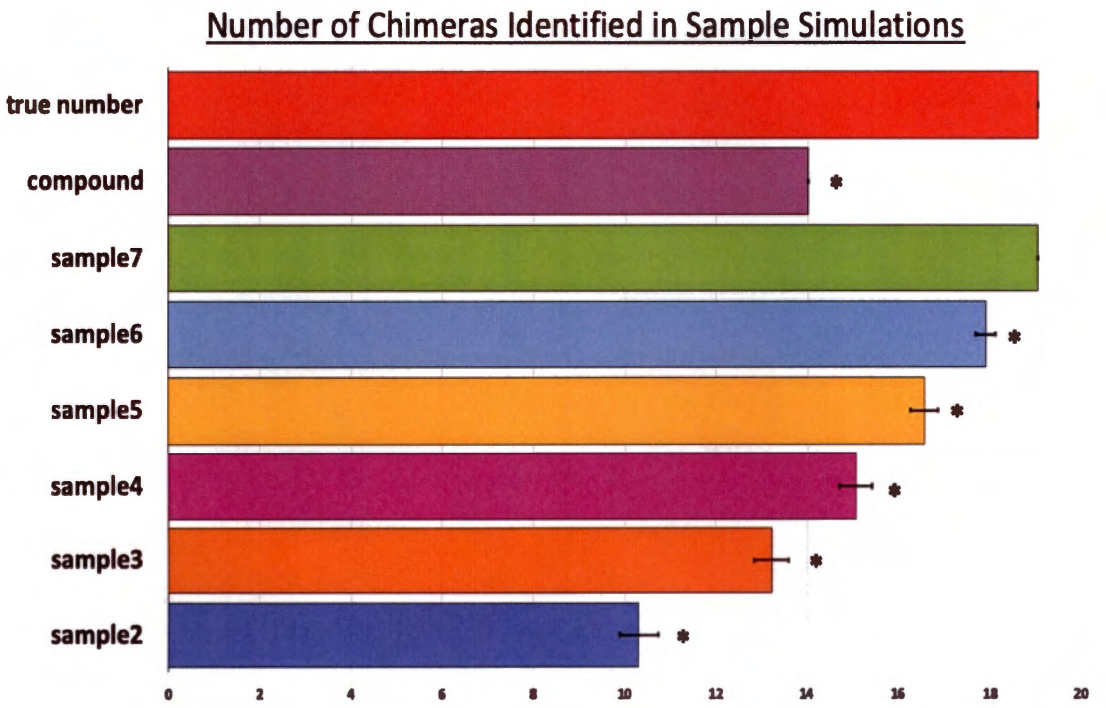


Figure 9c. Number of chimeras detected by simulated sampling method. Values shown for sample simulations are mean values over 1000 bootstrap replicates with error bars showing 95% confidence intervals. The top row shows the true number of chimeras (19). Bars shown with an asterisk were found to be significantly different from the true number of chimeras when compared with a paired one-tailed t-test ($p << 0.01$)



Appendix I: R code used for simulations of samples genotypes, calculation of clonal indices, statistical analyses, and graph creation

```
library("RClone",
lib.loc="/Library/Frameworks/R.framework/Versions/3.5/Resources/library")

library("plyr",
lib.loc="/Library/Frameworks/R.framework/Versions/3.5/Resources/library")

library("genepop"), library("poppr"), library("readr"), library("stringr")

library("stringi"), library("stats"), library("splitstackshape")

AllCols <- read_csv("AllNoSep_ColsandPops.csv",

                    col_types = cols(dvex03 = col_character(),

                                     dvex05 = col_character(), dvex12 = col_character(),

                                     dvex14 = col_character(), dvex20 = col_character(),

                                     dvex23 = col_character(), dvex26 = col_character(),

                                     dvex30 = col_character(), dvex32 = col_character(),

                                     dvex42 = col_character(), x = col_number(),

                                     y = col_number()))


UMP<-AllCols[105:nrow(AllCols),1:ncol(AllCols)]

HMB<-AllCols[1:104,1:ncol(AllCols)]

####name variables, get population and sample data, split into 2 columns, then SORT####

#combined datasets:

dat<-AllCols

alleles<-
data.frame(cSplit(dat[,6:15],c("dvex03","dvex05","dvex12","dvex14","dvex20","dvex23","dve
x26","dvex30","dvex32","dvex42")))

#popvec/colvec for all:

colvec<-c(split(dat[,2],"/n"))

colvec<-unlist(colvec,recursive=TRUE,use.names=FALSE)
```

```
popvec<-c(split(dat[,3],"/n"))

ALL_dvexGC<-convert_GC(alleles,3)

sort_all(ALL_dvexGC)

#For HMB:

dat1<-HMB

alleles1<data.frame(cSplit(dat1[,6:15],c("dvex03","dvex05","dvex12","dvex14","dvex20","dvex23","dvex26","dvex30","dvex32","dvex42")))

HMBdvexGC<-convert_GC(alleles1,3)


#For UMP:

dat2<-UMP

alleles2<data.frame(cSplit(dat2[,6:15],c("dvex03","dvex05","dvex12","dvex14","dvex20","dvex23","dvex26","dvex30","dvex32","dvex42")))

UMPdvexGC<-convert_GC(alleles2,3)

#clonal indices

HMB_clonalindex<-clonal_index(HMBdvexGC)

UMP_clonalindex<-clonal_index(UMPdvexGC)


CI_ALL_cols<-clonal_index(ALL_dvexGC,vecpop=colvec)

CI_ALL<-clonal_index(ALL_dvexGC,vecpop=NULL)

HMBUMPCI<-rbind(HMB_clonalindex, UMP_clonalindex)

write.csv(HMBUMPCI,"ClonalIndex_7.24.19.csv")

#use this to subset chimeras:

chimeras<-subset(CI_ALL_cols,CI_ALL_cols$G>1)

HowManyChimeras<-nrow(chimeras)

#test for marker reliability with sample_loci:

lox<-sample_loci(ALL_dvexGC, vecpop=NULL, nbrepeat=100, He=TRUE, graph=TRUE, bar=TRUE)
```

```
#accumulation curves for units (uns) and loci (locs)

uns<sample_units(ALL_dvexGC,haploid=FALSE,vecpop=colvec,nbrepeat=1000,He=TRUE,grap
h=FALSE,bar=TRUE)

uns_chims<-uns

 E4_BS<-uns_chims$E4$raw_MLG

 E15_BS<-uns_chims$E15$raw_MLG

 FP_BS<-uns_chims$FP$raw_MLG

 H0_BS<-uns_chims$H0$raw_MLG

 Mc_BS<-uns_chims$Mc$raw_MLG

 U1_BS<-uns_chims$UMP1$raw_MLG

 U2_BS<-uns_chims$UMP2$raw_MLG

 U3_BS<-uns_chims$UMP3$raw_MLG

 U5_BS<-uns_chims$UMP5$raw_MLG

 U6_BS<-uns_chims$UMP6$raw_MLG

 U9_BS<-uns_chims$UMP9$raw_MLG

 U10_BS<-uns_chims$UMP10$raw_MLG

 U13_BS<-uns_chims$UMP13$raw_MLG

 U14_BS<-uns_chims$UMP14$raw_MLG

U15_BS<-uns_chims$UMP15$raw_MLG

 U17_BS<-uns_chims$UMP17$raw_MLG

 U18_BS<-uns_chims$UMP18$raw_MLG

 U19_BS<-uns_chims$UMP19$raw_MLG

 U20_BS<-uns_chims$UMP20$raw_MLG

#separate them by # of sample units

raw2<-data.frame(cbind( E4_BS[2],E15_BS[2],FP_BS[2],H0_BS[2],Mc_BS[2],
U1_BS[2],U2_BS[2],U3_BS[2],U5_BS[2],U6_BS[2],U9_BS[2],U10_BS[2],U13_BS[2],U14_BS[2],U15
_BS[2],U17_BS[2],U18_BS[2],U19_BS[2],U20_BS[2]))

chimeraset2<-apply(raw2, 1, function(x) length(x[x>1]))
```

```
set2labs<-as.factor(rep("2", length(chimeraset2)))

set2<-data.frame(cbind(rep("2", length(chimeraset2)),as.numeric(chimeraset2)))

raw3<-data.frame(cbind(E4_BS[3],E15_BS[3],FP_BS[3],H0_BS[3],Mc_BS[3],

U1_BS[3],U2_BS[3],U3_BS[3],U5_BS[3],U6_BS[3],U9_BS[3],U10_BS[3],U13_BS[3],U14_BS[3],U15
_BS[3],U17_BS[3],U18_BS[3],U19_BS[3],U20_BS[3]))

chimeraset3<-apply(raw3, 1, function(x) length(x[x>1]))

set3<-data.frame(cbind(rep("3", length(chimeraset3)),as.numeric(chimeraset3)))

raw4<-data.frame(cbind(E4_BS[4],E15_BS[4],FP_BS[4],H0_BS[4],Mc_BS[4],

U1_BS[4],U2_BS[4],U3_BS[4],U5_BS[4],U6_BS[4],U9_BS[4],U10_BS[4],U13_BS[4],U14_BS[4],U15
_BS[4],U17_BS[4],U18_BS[4],U19_BS[4],U20_BS[4]))

chimeraset4<-apply(raw4, 1, function(x) length(x[x>1]))

set4<-data.frame(cbind(rep("4", length(chimeraset4)),as.numeric(chimeraset4)))

raw5<-data.frame(cbind(

E4_BS[5],E15_BS[5],FP_BS[5],H0_BS[5],Mc_BS[5],

U1_BS[5],U2_BS[5],U3_BS[5],U5_BS[5],U6_BS[5],U9_BS[5],U10_BS[5],U13_BS[5],U14_BS[5],U15
_BS[5],U17_BS[5],U18_BS[5],U19_BS[5],U20_BS[5]))

chimeraset5<-apply(raw5, 1, function(x) length(x[x>1]))

set5<-data.frame(cbind(rep("5", length(chimeraset5)),as.numeric(chimeraset5)))

raw6<-data.frame(cbind(

E4_BS[6],E15_BS[6],FP_BS[6],H0_BS[6],Mc_BS[6],
U1_BS[6],U2_BS[6],U3_BS[6],U5_BS[6],U6_BS[6],U9_BS[6],U10_BS[6],U13_BS[6],U14_BS[6],U15
_BS[6],U17_BS[6],U18_BS[6],U19_BS[6],U20_BS[6]))

chimeraset6<-apply(raw6, 1, function(x) length(x[x>1]))

set6<-data.frame(cbind(rep("6", length(chimeraset6)),as.numeric(chimeraset6)))

raw7<-data.frame(cbind(E4_BS[7],E15_BS[7],FP_BS[7],H0_BS[7],Mc_BS[7],

U1_BS[7],U2_BS[7],U3_BS[7],U5_BS[7],U6_BS[7],U9_BS[7],U10_BS[7],U13_BS[7],U14_BS[7],U15
_BS[7],U17_BS[7],U18_BS[7],U19_BS[7],U20_BS[7]))

chimeraset7<-apply(raw7, 1, function(x) length(x[x>1]))

set7<-data.frame(cbind(rep("7", length(chimeraset7)),as.numeric(chimeraset7)))
```

```
#Put simulated # chimeras in df, get mean and SD

SimChisTot1<-data.frame(chimeraset2, chimeraset3, chimeraset4, chimeraset5, chimeraset6,
chimeraset7)

SimChisTot<-data.frame(rbind(set2, set3, set4, set5, set6, set7))

names(SimChisTot)<-c("SampleUnits", "ChimerasDetected")

SimChisTot$SampleUnits<-as.numeric(as.character(SimChisTot$SampleUnits))

SimChisTot$SampleUnits<-as.factor(SimChisTot$SampleUnits)

p2<-ggplot(data=SimChisTot, aes(x=SampleUnits,y=ChimerasDetected))

p2+geom_point()

p2+labs(x="number of units sampled",y="number of chimeras detected")


Sim1<-data.frame(

NumUnits<-c("2 units","3 units","4 units","5 units","6 units","7 units"),

SimMeans<-c(apply(SimChisTot1,2,mean)),

SimSDs<-apply(SimChisTot1, 2, sd))


the_truth<-c(rep(19,1000))

ttest2<-t.test(chimeraset2, mu=19, alternative="two.sided", conf.level=0.95)

ttest3<-t.test(chimeraset3, mu=19, alternative="two.sided", conf.level=0.95)

ttest4<-t.test(chimeraset4, mu=19, alternative="two.sided", conf.level=0.95)

ttest5<-t.test(chimeraset5, mu=19, alternative="two.sided", conf.level=0.95)

ttest6<-t.test(chimeraset6, mu=19, alternative="two.sided", conf.level=0.95)

ttest7<-t.test(chimeraset7, mu=19, alternative="two.sided", conf.level=0.95)

SimSEs<-SimSDs/sqrt(20)

Minerror<-c(SimMeans-SimSEs)

Maxerror<-c(SimMeans+SimSEs)
```

```
Chis_and_error<-data.frame(SimMeans,SimSDs,Minerror,Maxerror)


b<-ggplot(data=Sim1,aes(x=NumUnits, y=SimMeans))

b+geom_bar(stat="identity")+

 geom_errorbar(ymin=Minerror,ymax=Maxerror)+

 labs(title="Simulated Sampling Units vs. Number of Chimeras ", x="Sample Units",
y="Chimeric Colonies Detected",

    size=10)

 call<-lm(SimMeans~NumUnits, data=Sim1)

chigenos<-c(chimeras$G)


chigenos<-
c(rep(2,49),rep(5,7),rep(3,7),rep(6,7),rep(2,7),rep(5,7),rep(3,7),rep(4,7),rep(4,7),rep(5,7),re
p(3,7),rep(2,14))


###Get the residuals for these calculations....


ColNames_Chimeras<-c(rep("E4",7),rep("E15",7),rep("FP",7),rep("H0",7),rep("MC",7),
rep("UMP1",7),rep("UMP2",7),rep("UMP3",7),rep("UMP5",7),rep("UMP6",7),rep("UMP9",7),
rep("UMP10",7),rep("UMP13",7),rep("UMP14",7),rep("UMP15",7),rep("UMP17",7),rep("UMP
18",7),rep("UMP19",7),rep("UMP20",7))

chigenos<-
c(rep(2,49),rep(5,7),rep(3,7),rep(6,7),rep(2,7),rep(5,7),rep(3,7),rep(4,7),rep(4,7),rep(5,7),re
p(3,7),rep(2,14))

table_of_residuals<-rbind(

resE4_BS<-uns_chims$E4$res_MLG, resE15_BS<-uns_chims$E15$res_MLG,

resFP_BS<-uns_chims$FP$res_MLG, resH0_BS<-uns_chims$H0$res_MLG, resMc_BS<-
uns_chims$Mc$res_MLG, resU1_BS<-uns_chims$UMP1$res_MLG, resU2_BS<-
uns_chims$UMP2$res_MLG, resU3_BS<-uns_chims$UMP3$res_MLG, resU5_BS<-
uns_chims$UMP5$res_MLG, resU6_BS<-uns_chims$UMP6$res_MLG, resU9_BS<-
uns_chims$UMP9$res_MLG, resU10_BS<-uns_chims$UMP10$res_MLG, resU13_BS<-
uns_chims$UMP13$res_MLG, resU14_BS<-uns_chims$UMP14$res_MLG, resU15_BS<-
uns_chims$UMP15$res_MLG, resU17_BS<-uns_chims$UMP17$res_MLG, resU18_BS<-
```

```
uns_chims$UMP18$res_MLG, resU19_BS<-uns_chims$UMP19$res_MLG, resU20_BS<-
uns_chims$UMP20$res_MLG)

table_of_residuals<-data.frame(cbind(ColNames_Chimeras,table_of_residuals))

proportionMLGS<-table_of_residuals$mean_MLG/chigenos

colors<-rep("color",length(colnames))

UnitsVsGenotypes<data.frame(cbind(ColNames_Chimeras,as.numeric(table_of_residuals$nb_
units),as.numeric(table_of_residuals$mean_MLG),as.numeric(chigenos),as.numeric(proportio
nMLGS),colors))

names(UnitsVsGenotypes)<-
c("Colony","Units_Sampled","Mean_MLGs","True_MLGs","Percent_True_MLGs","color")

UnitsVsGenotypes<-subset(UnitsVsGenotypes, UnitsVsGenotypes$Units_Sampled!=1)

call2<-lm(as.numeric(Units_Sampled)~Mean_MLGs, data=UnitsVsGenotypes)

#split up chimeras into different number of genotypes

bichimeras<-subset(UnitsVsGenotypes, UnitsVsGenotypes$True_MLGs==2)

bichimeras$color="purple"

trichimeras<-subset(UnitsVsGenotypes, UnitsVsGenotypes$True_MLGs==3)

trichimeras$color="red"

tetrachimeras<-subset(UnitsVsGenotypes, UnitsVsGenotypes$True_MLGs==4)

tetrachimeras$color="blue"

pentachimeras<-subset(UnitsVsGenotypes, UnitsVsGenotypes$True_MLGs==5)

pentachimeras$color="magenta"

hexachimera<-subset(UnitsVsGenotypes, UnitsVsGenotypes$True_MLGs==6)

hexachimera$color="hotpink"


chimeras_by_number<-rbind(bichimeras, trichimeras, tetrachimeras, pentachimeras,
hexachimera)

chimeras_by_number$Mean_MLGs<-
as.numeric(as.character(chimeras_by_number$Mean_MLGs))

chimeras_by_number$Percent_True_MLGs<-
as.numeric(as.character(chimeras_by_number$Percent_True_MLGs))
```

```
chimeras_by_number$Units_Sampled<-
as.numeric(as.character(chimeras_by_number$Units_Sampled))

####LINEAR REGRESSION UNITS SAMPLED VS. PERCENT TRUE MLGS DETECTED

call2.2<-lm(as.numeric(bichimeras$Units_Sampled)~bichimeras$Percent_True_MLGs,
data=bichimeras)

call3<-lm(as.numeric(trichimeras$Units_Sampled)~trichimeras$Percent_True_MLGs,
data=trichimeras)

call4<-lm(as.numeric(tetrachimeras$Units_Sampled)~tetrachimeras$Percent_True_MLGs,
data=tetrachimeras)

call5<-lm(as.numeric(pentachimeras$Units_Sampled)~pentachimeras$Percent_True_MLGs,
data=pentachimeras)

call6<-lm(as.numeric(hexachimera$Units_Sampled)~hexachimera$Percent_True_MLGs,
data=hexachimera)

stats_2g<-summary(call2.2)

stats_3g<-summary(call3)

stats_4g<-summary(call4)

stats_5g<-summary(call5)

stats_6g<-summary(call6)

unitsVsMLGs<-
boxplot(table_of_residuals$mean_MLG~table_of_residuals$nb_units,xlab="Number of units
sampled",ylab="Mean number of MLGs found")

unitsVsMLGs2<-boxplot(proportionMLGS~table_of_residuals$nb_units,xlab="Number of units
sampled",ylab="Proportion of MLGs found")

p<-ggplot(data=SimChisTot) +geom_violin(aes())

p+labs(x="number of units sampled",y="number of chimeras detected")

p2<-ggplot(data=table_of_residuals,
aes(x=as.factor(table_of_residuals$nb_units),y=proportionMLGS)) +geom_violin()

p2+labs(x="number of units sampled",y="proportion of MLGs detected")

p3<-ggplot(data=chimeras_by_number, aes(x=chimeras_by_number$Units_Sampled,
y=chimeras_by_number$Mean_MLGs))

p3+geom_point(aes(colour=factor(chimeras_by_number$True_MLGs))) +
```

```
labs(x="simulated sample size", y="mean number of multilocus genotypes detected",title=
"Multilocus Genotypes in Simulated Samples", color="True number of MLGs") +

theme(legend.position = "bottom",legend.title = element_text(size=9, color="black"),
legend.title.align = 0.5,

    legend.background = element_rect(fill="lightblue", size=0.5, linetype="solid", colour
="darkblue"),

axis.text.y = element_text(size=10, vjust=0.5), plot.title = element_text(vjust=0.5))+

coord_equal()


p4<-ggplot(data=chimeras_by_number, aes(x=chimeras_by_number$Units_Sampled,
y=chimeras_by_number$Percent_True_MLGs))

p4+geom_point(aes(colour=chimeras_by_number$True_MLGs,
shape=chimeras_by_number$True_MLGs, alpha=0.9))  +

 labs(x="simulated sample size", y="percent of true multilocus genotypes detected")+

  theme(legend.position = "bottom",legend.title = element_text(size=8, color="black"),

     legend.title.align = 0,

     legend.background = element_rect(fill="lightblue", size=0.5, linetype="solid", colour
="darkblue"),

     axis.text.y = element_text(size=10, vjust=0.5), plot.title = element_text(vjust=0.5))+

 scale_y_continuous()


#poppr

motif_lengths<-c(3,2,2,2,2,4,2,2,3,2)

GenCloneDat<-read.genalex("UMP_HMB_ALL_6.16.19.csv")

GenCloneDat1<-as.genclone(GenCloneDat)

divtab<-diversity_ci(GenCloneDat1, ci=95, rarefy=TRUE, plot=TRUE)

dm1<-diss.dist(GenCloneDat1)


popcolsHMB<-read.genalex("HMBCols_GenAlEx.csv")
```

```
popcolsUMP<-read.genalex("template_UMP.csv")


dm2<-diss.dist(popcolsHMB)

dm3<-diss.dist(popcolsUMP)

minspanUMP<-poppr.msn(popcolsUMP, dm3)

minspanHMB<-poppr.msn(popcolsHMB, dm2)

network1<-imsn(GenCloneDat1, minspan, wscale=TRUE, showplot=TRUE,nodelab=1)

sexval<-psex(GenCloneDat1)

min(sexval)

max(sexval)

PopGenTab<-poppr(GenCloneDat,clonecorrect=FALSE, legend=TRUE, sample=999,
plot=TRUE)


#statistical analyses

Simsds<-c(0,0,1.86,1.82,1.49,1.33,1.01,0)

SimSEs<-Simsds/sqrt(1000)


#95% confidence interval for each:

#use qt function b/c data are not normally distributed

CIs<-qt(0.975,df=1000)*SimSEs

CIs.upper1<-vals+CIs

CIs.lower1<-vals-CIs

Tot1<-data.frame(cbind(methods,vals, CIs.upper1, CIs.lower1))


####test statistics for number of chimeras:

Xmeras <-

 matrix(c(10, 5, 6, 14),
```

```
    nrow = 2,
    dimnames = list("HMB" = c("homogeneous", "chimeric"),
            "Umpqua" = c("homogeneous", "chimeric")))
Xmeras
p<-c(0.33, 0.7)
mcnemar.test(Xmeras)
x<-c(10,5)
y<-c(6,14)
truchis<-c(19,0)
compoundchis<-c(14,5)
sim2<-c(10,9)
sim3<-c(14,5)
sim4<-c(15,4)
sim5<-c(17,2)
sim6<-c(18,1)
sim7<-c(19,0)
signTest(truchis, alternative = "greater", mu = 0, paired = TRUE,
    conf.level = 0.9)

###binomial tests for significance in chimera detection 95%ci
btcomp<-binom.test(compoundchis, n, p = 1,
          alternative = c("less"),
          conf.level = 0.95)

bt2<-binom.test(sim2, n, p = 1,
     alternative = c("less"),
```

```
        conf.level = 0.95)

bt3<-binom.test(sim3, n, p = 1,
        alternative = c("less"),
        conf.level = 0.95)
bt4<-binom.test(sim4, n, p = 1,
           alternative = c("less"),
           conf.level = 0.95)

bt5<-binom.test(sim5, n, p = 1,
           alternative = c("less"),
           conf.level = 0.95)

bt6<-binom.test(sim6, n, p = 1,
        alternative = c("less"),
        conf.level = 0.95)

bt7<-binom.test(sim7, n, p = 1,
           alternative = c("less"),
           conf.level = 0.95)
binomialresults<-c(btcomp,bt2,bt3,bt4,bt5,bt6,bt7)

#testing for differences in the two populations
prop.test(Xmeras, p=p, alternative="greater", conf.level = 0.95, correct=FALSE)
```