Genetic Population Structure in Blue Crabs (Callinectes sapidus): High Resolution Population Genomics of a High Gene Flow Species

Joseph Neigel[1], Sophie Plouviez[1], Timothy Sullivan[2], Bree Yednock[3], and Vanessa Rashbrook[4]

[1]University of Louisiana at Lafayette
[2]Gloucester Marine Genomics Institute
[3]South Slough National Estuarine Research Reserve
[4]University of California Davis

May 5, 2020

Abstract

Widespread dispersal of progeny is expected to result in enough gene flow to maintain genetic homogeneity over large areas. Surveys of genetic markers in species with planktonic larvae have mostly confirmed this expectation. However, genetic structure has occasionally been found at small spatiotemporal scales and interpreted as evidence of restricted dispersal, natal homing, sweepstakes reproductive success, or natural selection. We investigated genetic population structure in blue crabs from the Atlantic and Gulf of Mexico coasts of North America. Sampling was most intensive from five estuaries along the coast of Louisiana, with megalopae, juveniles and adults sampled from 2010 to 2016. 1446 individuals were genotyped at 2486 SNPs in 1363 putative protein-coding loci. Levels of differentiation between locations were consistently low, but significant differentiation was found among locations and among years. No evidence was found for chaotic genetic patchiness or sweepstakes reproductive success: no genetic differentiation was detected among collections of megalopae and none of the sampled individuals were closely related. Our results indicate that gene flow in blue crabs maintains near genetic homogeneity from the northern Gulf of Mexico through the Atlantic coast of North America.

1 | INTRODUCTION

Gene flow from planktonic larval dispersal has the potential to maintain genetic panmixia over large distances in the marine realm (Shaklee & Bentzen, 1998), although this potential is not always realized. Possible causes of departures from panmixia include historical biogeographic separation, physical oceanographic barriers, larval behavior, sweepstakes reproductive success (SRS), and natural selection (reviewed in Hellberg, 2009). Disentangling these causes is challenging because they can act synergistically. Thus, a genetic break formed by a vicariant event could subsequently be maintained by selection (Schneider-Broussard, Felder, Chlan, & Neigel, 1998), or larval behavior could exploit oceanographic features to limit dispersal (Kingsford et al., 2002). Sometimes it is possible to exclude mechanisms a priori, thus in species with low fecundity SRS can be eliminated as the cause of chaotic patchiness (Cornwell, Fisher, Morgan, & Neigel, 2016). More generally, distinctive patterns of genetic population structure are considered signatures of specific mechanisms (Hellberg, Burton, Neigel, & Palumbi, 2002). Extensive sampling is needed to adequately characterize these patterns, which can appear at different scales across time, space and life stage. Furthermore, numerous marker loci are needed to detect and quantify the typically low levels of population structure in marine populations (Waples, 1998). With the above considerations in mind, we undertook a survey of single nucleotide

polymorphisms (SNPs) in the blue crab, *Callinectes sapidus* , an economically important, well-studied species (Kennedy & Cronin, 2007).

The life cycle of the blue crab favors dispersal away from natal habitats (Epifanio, 2007). Females release planktonic larvae offshore where they are transported by currents for 4-6 weeks. Megalopae (post-larvae) migrate to coastal brackish water habitats where they settle and metamorphose into juveniles. After mating, mature females migrate to offshore spawning sites, sometimes travelling hundreds of kilometers (Aguilar et al., 2005; Gelpi, Fry, Condrey, Fleeger, & Dubois, 2013). Most surveys of genetic variation in North American blue crabs have found no detectable genetic differentiation (Berthelemy-Okazaki & Okazaki, 1997; Lacerda et al., 2016; McMillen-Jackson & Bert, 2004; Yednock & Neigel, 2014) or very little (McMillen-Jackson, Bert, & Steele, 1994; Plough, 2017) over distances up to thousands of kilometers. In the surveys that found slight differentiation, it was detected near the limits of their statistical power (discussed in Yednock & Neigel, 2014), which suggests that additional genetic structure is present at undetectable levels. A few studies have found slight temporal or life-stage shifts in the genetic composition of blue crab populations (Feng, Williams, & Place, 2017; McMillen-Jackson et al., 1994; Yednock & Neigel, 2014). Temporal shifts can cause chaotic genetic patchiness, a pattern of fluctuating small-scale spatial variation first described in limpets (Johnson & Black, 1982, 1984). Kordos and Burton (1993) reported anomalously high levels of temporal and spatial differentiation for allozyme loci in blue crabs from the coast of Texas. It is possible that pronounced genetic structure such as they described develops episodically or at specific locations in blue crab populations, but see Sullivan and Neigel (2017) on the misidentification of specimens as a plausible explanation for some of these anomalous findings.

The native range of *C. sapidus* is disjunct along the Atlantic coast of the Americas and encompasses at least two distinct genetic units. The northern portion of the range extends from Massachusetts in the U.S. to Venezuela while the southern portion extends from Bahia, Brazil to Mar del Plata in Argentina (Santos & D'Incao, 2004). Yednock and Neigel (2014) found that blue crabs from the northern Gulf of Mexico were genetically different from those from Venezuela (F_{ST} for each locus between 0.06 and 0.68) but not from Mexico. Rodrigues and co-workers (2017) found two mitochondrial lineages in blue crabs, one in both portions of the species' range and the other exclusive to the southern portion. It is likely that unintended sampling of multiple species of *Callinectes* has been a source of error in blue crab ecological and population genetic research, especially for studies that included early life stages that are difficult to identify to species (Sullivan and Neigel, 2017).

Our sampling program was designed to be sensitive to both low levels of genetic differentiation and episodic genetic structure. We sampled individuals from multiple locations, time points, and life stages to characterize patterns of genetic differentiation rather than simply test for its presence. Some SNPs were in genes targeted for their presumed functions. These criteria led us to a genotyping method, the Infinium Assay (Illumina), which is seldom used for population genetics of non-model organisms. We were able to use the Infinium assay because an annotated transcriptome allowed us to identify SNPs in protein-coding regions and a draft genome helped us locate and avoid introns.

Our findings support the main conclusions of most previous studies: we detected little genetic population structure in blue crabs from the Gulf of Mexico and North Atlantic. In addition, there was no evidence of SRS: we found neither chaotic genetic patchiness or closely related individuals. Overall, it appears that levels of gene flow are high among North American blue crabs as expected from their biology. Although these findings might appear unremarkable, they point to an opportunity to observe the effects of natural selection on genetic structure against a background that is relatively free of other structuring mechanisms. Here we focus on detection of genetic differentiation in blue crabs and its plausible causes, exclusive of natural selection.

2 | METHODS

2.1 | Specimen Collection

Blue crabs were sampled from the northern Gulf of Mexico, the U.S. Atlantic coast and Caribbean (Fig. 1A, Table 1). Sampling was concentrated at five estuarine locations (FWC, GIL, LUM, RBC and RWR) along the Louisiana coast (Fig. 1B) from which 1234 adults, juveniles and megalopae were sampled over a 6-year period and genotyped (Table S1). Additional samples from more distant locations were generously provided by colleagues. Adults and juveniles were collected with baited crab pots, hoop nets, and hand lines. Carapace width from point to point was measured to the nearest millimeter and sex was determined by the shape of the abdominal tergites. Males larger than 120 mm and females with dome-shaped abdominal tergites were considered mature (following Newcombe, Sandoz, & Rogers-Talbert, 1949; Van Engel, 1990). The second walking leg on the right side of each crab was removed and stored in 95% ethanol at 4@C. Megalopae were sampled with modified "hogs-hair" collectors (Metcalf, van Montfrans, Lipcius, & Orth, 1995) deployed for approximately 24 hours, those visually identified as *C. sapidus* were removed, washed with ambient seawater and preserved in pre-chilled 95% ethanol at 4@C.

2.2 | DNA Extraction

DNA was extracted from approximately 2 mg of muscle tissue from each adult or juvenile, or from entire megalopae using Nucleospin® Tissue Kits (Machery-Nagel) in an epMotion 5075 TMX liquid handling workstation (Eppendorf) following the manufacturer's protocols. DNA concentrations were measured with a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific) and adjusted to between 50 and 150 ng/µl in a volume of at least 15 µl by either concentration in a SpeedVac vacuum concentrator (Thermo Fisher Scientific) or dilution with deionized water.

2.3 | Selection of SNPs for Genotyping

We selected 2958 SNPs for genotyping on a custom 24-chip Infinium® iSelect® array (Illumina) with 5373 bead types. SNPs were selected for assessment of data quality, analysis of population structure and testing for natural selection. Gene annotation and SNP discovery were based on an assembled transcriptome representing seven individuals (Yednock, Sullivan, & Neigel, 2015). Statistical power to detect genetic differentiation increases with the number of alleles (Ryman et al., 2006), so it should be advantageous to combine alleles of neighboring SNPs into greater numbers of haplotypes. To this end, we selected multiple linked SNPs for some loci. We considered 272 gene loci in which SNPs were assayed as candidates for selection by environmental stress or pathogens. These loci had been assigned the gene ontology terms: *cold* , *defense* , *detox* , *ecdys* ,*heat* , *hypoxia* , *osmo* or *xenobio* by Yednock and co-workers (2015).

2.4 | Design of the Infinium Bead Array

The Infinium assay is based on the specificity of single-base extensions of DNA duplexes formed by hybridization of bead-anchored DNA probes to genomic target sequences. Based on the transcriptomes from Yednock and co-workers (2015), SNPs were required to be biallelic, have a minor allele frequency of at least 0.3, a high-quality read depth of at least 50, and be flanked by a region of at least 60 bp without other SNPs. Introns were identified by comparing transcriptomic sequences with genomic sequences identified by BLAST searches (Altschul, Gish, Miller, Myers, & Lipman, 1990) of a draft genome of *C. sapidus* (generously provided by T. Schultz). Probes that overlapped introns were dropped from further consideration. Probe designs were evaluated with the Illumina online Assay Design Tool; only probes with a final score greater than 0.7 were considered further. To evaluate the species-specificity of the SNP assays, a representative of a congener (*Callinectes similis*) was also genotyped. All genotyping was performed by the DNA Technologies Core at the University of California Davis Genome Center.

2.5 | Genotyping SNPs with Two Probes

Coding regions are more polymorphic in the blue crab genome (1 SNP per 13.3 bp; Yednock & Neigel, 2014) than in the human genome (1 SNP per 346 bp; Cargill et al., 1999) for which the Infinium assay has been extensively tested (Steemers et al., 2006). This was a concern, because polymorphisms within the probe-binding region can cause genotyping errors. To detect genotyping errors, improve call rates, and identify loci with null alleles we used two probes for 1750 of the 2958 genotyped SNPs (59%). One probe was designed to hybridize downstream of the SNP and the other upstream on the opposite strand. For each crab, if a genotype was not called for one of the probes the genotype called for the other probe was assigned. Genotypes were scored as unknown if different genotypes were called for the two probes or no call was made for either probe. A probe with null alleles would have lower apparent heterozygosity than an unaffected probe for the same locus. The deviation from expected heterozygosity was calculated as $(H_{observed} - H_{expected}) / H_{expected}$ for both probes; if their values differed by more than 0.05 that locus was not used for population structure analysis.

2.6 | Genotype Calling with GenomeStudio

Initially, 2958 SNPs were genotyped from 1508 putative loci. We used 665 probes for A/T and C/G SNPs, and 4043 probes for A/G, A/C, T/G, and T/C SNPs, requiring 5373 bead types. SNP genotypes were inferred by the GenTrain 3.0 algorithm in the Genotyping Module of GenomeStudio 2.0.2 (Illumina), which also calculates a GenCall score (the frequency of reliable calls). Data from probes with GenCall scores below 0.85 were discarded. For population structure analysis, only crabs for which 85% of loci were successfully genotyped were included. Scatter plots of clusters of genotypes were visually assessed for each locus; genotypes that did not belong to a well-defined cluster were scored as unknown.

2.7 | Illumina Infinium Genotyping versus Sequencing

Yednock and Neigel (2014) sequenced portions of protein-coding genes in blue crabs. 176 of these crabs were also genotyped in the present study, allowing a comparison between Infinium genotyping and sequencing. Comparisons were made for two SNPs in *ATP/ADP translocase* (*ant*), two in*ATP-synthase subunit 9* (*atps*), and three in*trehalose 6-phosphate synthase* (*tps*). Deviation from Hardy-Weinberg proportions (F) determined both from sequences and from Illumina genotypes were estimated with Genepop 4.2 (Rousset, 2008).

2.8 | Phasing SNPs in Multi-SNP Loci

SNPs within the same locus were phased and combined into haplotypes with PHASE 2.1 (Stephens & Donnelly, 2003; Stephens, Smith, & Donnelly, 2001). Initially, 5 replicate runs of 1000 iterations with a thinning interval of 1 and a burn-in of 1000 were used. If replicates differed in haplotype assignments, 5 more were run with 10,000 iterations, and if those differed, 5 were run with 50,000 iterations and burn-ins of 5,000. If less than 90% of individuals were each assigned haplotypes with at least 90% probability after PHASE runs, progressively smaller subsets of SNPs were evaluated until a subset was found that met this criterion.

2.9 | Tests for Linkage Disequilibrium

After SNPs within the same locus were phased and combined into haplotypes, tests for linkage disequilibrium (LD) between pairs of loci were run in Genepop 4.7.0. Genepop evaluates the significance of deviations from expected proportions of 2-locus diploid genotypes. Initial runs combined all genotyped individuals of *C. sapidus*into a single population sample and used default MCMC parameters of 10,000 dememorization

steps, batch lengths of 5000 and 100 batches. As recommended by Waples (2014), the distribution of p -values for each test was compared with the expected uniform distribution. Loci with the lowest and highest p -values were retested in individual populations to assess if LD was due to a Wahlund effect (Sinnock, 1975).

2.10 | Analysis of Population Structure

We began by calculating F_{ST} and testing for the presence of population structure among locations, years and life stages. When significant overall heterogeneity was found, two approaches were used to characterize the pattern of heterogeneity: 1) tests of heterogeneity in defined subsets of samples, with sequential Bonferroni adjustments of the Type I error rate, α, which was set to 0.05, and 2) pairwise estimates of F_{ST} between sampling units, with Bonferroni adjustments of α and control of the False Discovery Rate (FDR) (Benjamini & Hochberg, 1995) applied to tests of $F_{ST} > 0$. F_{ST} was estimated with Genepop 4.7.0. Tests for overall heterogeneity were conducted with CHIFISH (Ryman, 2006), which provides two estimates of p : one from χ^2 summed over loci and the other with Fisher's method of combining multiple independent tests. The significance of pairwise F_{ST} estimates ($F_{ST} > 0$) was determined with Arlequin (Schneider, Roessli, & Excoffier, 2000) using 10,000 permutations of the data.

Statistical power of tests for population differentiation was assessed with the Powsim_b version of POWSIM (Ryman & Palm, 2006), downloaded from http://internt.zoologi.su.se/~ryman/. Genetic drift was simulated for populations with effective sizes of 10,000 for 1 or 2 generations, with expected F_{ST} reaching 0.00005 and 0.0001 respectively. 1000 replicates were used to estimate statistical power and Type I error rates. Markov chain parameters for estimation of p by Fisher's method were 1000 dememorizations, 100 batches and 1000 iterations.

2.11 | Estimation of Coancestry

The coefficient of relationship (r) between individuals was estimated by the likelihood method of Milligan (2003) using Coancestry 1.0.1.8 (Wang, 2011), which requires estimates of population allele frequencies. We examined how estimates of r were affected by partitioning of the full sample for estimates of allele frequencies by life stage, sampling year, or sampling location.

3 | RESULTS

3.1 | Genotyping Call Rates and Reproducibility

A total of 2958 SNPs were genotyped; 910 of 1208 (75.3 %) genotyped with single probes passed quality assessment in GenomeStudio 2.0.2. For the 1750 SNPs genotyped with two probes, 1118 SNPs (63.9%) passed quality assessment for both probes, 470 (26.9%) for one probe, and 162 (9.3%) for neither probe. Overall, 2498 SNPs (84.4%) passed quality assessment for at least one probe and with two probes a success rate of 90.7% was achieved. After loci with significant deviations from Hardy-Weinberg proportions were removed, data for 2486 SNP loci were retained for population genetic analysis.

After probes with low call rates were eliminated the average call rate for probes was 99.0% and only two individuals (0.13%) had probe call rates below 70%. 1475 individuals (96.0%) had probe call rates meeting our cutoff (85%), 1447 (94.2%) had call rates of at least 99%, and 139 (9.0%) had call rates of 100%. Reproducibility of genotypes at each locus for individual crabs was nearly perfect. Of the 17 individuals that were genotyped twice, data for one was discarded because the call rate was below 84%. Among the remaining 16 individuals, identical SNP genotypes (across all loci) were called for both replicates of six individuals, and a genotype was called at one or two loci in one replicate but not the other for 10 individuals. There were no instances in which different genotypes were called for replicates of the same individual. After crabs

with low call rates or anomalous heterozygosity were removed from the data, genotype data was available for 1434 crabs.

3.2 | Genotyping and Non-target Species

Low call rates for non-target species are expected because of mismatches with probes designed for the target species (*C. sapidus*). Consistent with this expectation, the probe call rate for *Callinectes similis* (0513LUMCsimJ1) was only 77%. A megalopa collected near Galveston, Texas (GAL815_M45) with a probe call rate of 87% was subsequently identified by its mitochondrial 16S sequence as *C. rathbunae* (Fig. 2). Seven other megalopae from Louisiana (14RWR7516, FWC15M20, FWC15M21, FWC15M22, FWC15M23, FWC15M25, and FWC15M27) had probe call rates and SNP heterozygosities that clustered tightly with those of GAL815_M45 (Fig. 2). Comparisons based on which probes were unscorable and which allele was present at each homozygous SNP confirmed that these megalopae were genetically very similar to GAL815_M45 (Table S2). Between 93.3% and 96.5% of the 255 probes that were unscorable in GAL815_M45 were also unscorable in these megalopae, and between 97.8% and 98.5% of the 2220 SNP loci that were homozygous in GAL815_M45 were homozygous for the same allele in these megalopae. In contrast, the individual of *C. similis* (0513LUMCsimJ1) matched GAL815_M45 for only 46.6% of unscorable probes and 64.7% of alleles at homozygous loci.

Five specimens from Venezuela had low heterozygosities (near 0.1) and probe call rates between 88.2% and 99.5% (Fig. 2). They were similar in which alleles were present at homozygous loci, but not in which probes were unscorable and they did not share high percentages of either unscorable or homozygous loci with representatives of *C. rathbunae* or *C. similis* (Table S2). These Venezuelan specimens thus appear to represent a population lacking polymorphism at loci that are polymorphic in North American populations. Probe call rates were not exceptionally low for three of these individuals, while low probe call rates for the other two, which were museum specimens, may have been a consequence of inadequate preservation.

A single specimen from Louisiana (0813LUMCSapM017) had a low proportion of scorable loci but a uniquely high proportion of heterozygous loci (Fig. 2). We devised a test to determine if this individual could be a hybrid between *C. sapidus* and a crab with a genotype matching any of the low-heterozygosity individuals described above. We compared the genotype at each of the 977 homozygous loci in 0813LUMCSapM017 with the genotype for the same locus in each potential parental genotype (GAL815_M45, 0513LUMCsimJ1, 0000EMVCsapA1, 501EMVCsapA7, 0901ZLVCsapU1, 0901ZLVCsapU2, and 1099ZLVCsapU3). Following Mendelian principles, we should not observe cases in which a parental genotype is homozygous for an allele not present in its hybrid offspring. By this criterion, none of the genotypes we tested could be the parent of 0813LUMCSapM017 (Table S3). The number of loci that failed to meet this criterion ranged from 104 (9.8%) to 210 (18.8%).

3.4 | Illumina Infinium Genotyping versus Sequencing

Genotypes were successfully called by the Infinium assay for all 7 SNPs among the 176 blue crabs from which sequence data was available from Yednock and Neigel (2014). Sequences were available for *ATP/ADP translocase* (*ant*) from 149 crabs and for both *ATP-synthase subunit 9* (*atps*) and *trehalose 6-phosphate synthase* (*tps*) from 167 crabs. One of the SNPs in *tps* was monomorphic in this set of individuals and is not considered further. Overall, 94% of the 966 SNP genotypes determined by Infinium matched the sequences. However, the percent in agreement ranged from 76.5% for a SNP in *ant* to 100% for the two SNPs in *tps* . Yednock and Neigel (2014) reported a significant heterozygote deficiency for the *ant* locus and suggested this was due to null alleles. This interpretation is supported by Infinium genotyping: all 53 of the discrepancies between *ant* sequences and Infinium genotypes were instances of SNPs called as heterozygotes by Infinium appearing as homozygotes in sequences. Furthermore, departures from expected heterozygosity were large and highly significant for both SNPs in *ant* when genotyped by sequencing (F values of 0.404 and 0.389, p = 0.00 for both SNPs) but were not when genotyped by Infinium (F values of 0.096 and -0.110). For the

two SNPs in *atps* , there were a total of five discrepancies between sequences and Infinium genotypes and all except one were instances of genotypes called heterozygotes by the Infinium assay appearing as homozygotes in sequences.

3.5 | Phasing Haplotypes at Multi-SNP Loci

Two or more SNPS were successfully genotyped for 407 loci. After PHASE runs were completed, 211 (52%) of the loci met our 0.9/90% criteria (haplotypes inferred with at least 0.9 probability in at least 90% of individuals). Longer runs sometimes led to convergence among replicates but still failed to identify haplotype configurations that met the 0.9/90% criteria. For loci that did not meet these criteria we removed SNPs (collapsing the set of haplotypes) until the 0.9/90% criterion was satisfied for the haplotypes distinguished by the remaining SNPs. From an initial total of 1484 SNPs that were phased, data for 1095 SNPs (74.8 %) were retained and distinguished a total of 2196 haplotypes. The conversion of 1484 SNPs into 2196 haplotypes increased the total number of degrees of freedom from 1077 to 1789, a 66% increase. Genotype and specimen collection data used for analyses in this paper are available on GRIIDC (Neigel, 2017).

3.6 Linkage Disequilibrium

With all individuals pooled and all polymorphic loci tested for linkage disequilibrium (LD) there was an unexpectedly high proportion of$p$ values near zero (Fig. S1): 3,269 tests out of a total of 911,925 had p values below 0.001, 3.6 times more than the 912 expected under the null hypothesis of no LD. We retested the 10 loci with the lowest p values and the 10 loci with the highest p values in subsamples representing each combination of sampling location and year for the main sampling locations on the Louisiana coast (Table S1). The two resulting distributions were similar and both lacked an excess of low $p$values (Fig. S2), suggesting that LD in the pooled samples could be a Wahlund effect (Sinnock, 1975; Waples, 2014). Peaks at $p = 0$ were also investigated and were found to be associated with contingency tables that had cells with low counts.

3.7 | Statistical Power to Detect Structure

POWSIM was used to estimate the statistical power of theχ^2 and Fisher's exact tests that are used by CHIFISH to detect genetic differentiation among populations. For samples from the five main locations in Louisiana (with all life stages combined), Type I error rates for the χ^2 test and Fisher's exact test were 0.037 and 0.077 respectively with all loci used, 0.055 for χ^2 and 0.078 for Fisher's with haplotypes of multi-SNP loci, and 0.058 for $\chi^2$and 0.091 for Fisher's with single-SNP loci. Using all markers, statistical power to detect even weakly differentiated populations ($F_{ST} = 0.0001$) was considerable. The proportion of replicate simulations runs in which differentiation was significant with α set to 0.05 was 0.90 for χ^2 and 0.92 for Fisher's exact test; with multi-SNP loci the proportions were 0.70 forχ^2 and 0.72 for Fisher's, and with single-SNP loci the proportions were 0.65 for χ^2 and 0.74 for Fisher's. With the level of differentiation reduced to$F_{ST} = 0.00005$, the power estimates were 0.54 forχ^2 and 0.47 for Fisher's.

3.8 | Large-Scale Genetic Population Structure

Overall geographic differentiation among 15 locations in the Atlantic and GOM was slight but highly significant ($F_{ST} = 0.0002$; $p = 0.00000$ for χ^2, 0.00017 for Fisher's). Estimates of pairwise-F_{ST} between locations were small (-0.0016 to 0.0054, mean 0.00065), with the highest between the samples from JAC and RRC (Table S4). 17 of the 105 pairwise comparisons were significantly different from zero with the False Discover Rate (FDR) set to 0.05, and all were comparisons with samples from either JAC or RRC. Following a Bonferroni correction for multiple comparisons, 7 of the 105 comparisons were significant, and all were comparisons with samples from JAC. Without the JAC sample, overall geographic differentiation was not significant ($F_{ST} = 0.0001$; $p = 0.32$ for χ^2, 0.071 for Fisher's).

F_{ST} estimated among sampling years, with locations and life stages pooled, was small but statistically significant ($F_{ST} = 0.0001$; $p = 0.012$ for χ^2, 0.031 for Fisher's). This raises the question of whether apparent genetic differentiation among locations could be caused by differences among the years in which locations were sampled. However, for samples collected in 2010, which included 13 of the 17 locations, this does not appear to be the case. The estimated overall F_{ST} among locations in 2010 was 0.0004 (higher than the overall estimate forF_{ST} among locations pooled across years) and highly significant ($p = 0.00001$ for χ^2, 0.0043 for Fisher's).

3.9 | Patterns of Differentiation on Louisiana Coast

Blue crab megalopae, juveniles and adults were sampled from five locations on the coast of Louisiana in six different years. The estimated overall F_{ST} among the five locations (life stages and years pooled) was 0.0000. F_{ST}among years (life stages and locations pooled) was estimated at 0.0001 and was statistically significant ($p = 0.025$ for χ^2, 0.016 for Fisher's). In pairwiseF_{ST} estimates among the six sampling years (Table S5), two of the 15 comparisons were significant after control of the FDR at 0.05: 2010 vs. 2016 ($F_{ST} = 0.0003$, $p = 0.0014$) and 2013 vs. 2015 ($F_{ST} = 0.0002$, $p = 0.0059$).

Sources of heterogeneity among collection years in Louisiana were identified by testing subsets of samples. First, heterogeneity among years was tested separately at each of the five locations; none of these were significant. Second, heterogeneity among years was tested separately for each of the three life-stages. Heterogeneity was significant among years for juveniles ($F_{ST} = 0.0003$; $p = 0.0059$ for χ^2, 0.013 for Fisher's,), but not for adults ($F_{ST} = 0.0002$), and was absent for megalopae (F_{ST} = -0.0001). For juveniles (Table S6), the only between-year F_{ST}pairwise estimate that was significantly greater than zero was between 2011 and 2013 ($F_{ST} = 0.0009$; $p = 0.0007$). Settling megalopae were collected from four locations (FWC, GIL, LUM and RWR) in at least five different years for a total of 21 collections. We found no evidence for heterogeneity among collections of megalopae (F_{ST} = -0.0004).

3.10 | Coancestry among Blue Crabs

Initial estimates of r , the coefficient of relationship, between pairs of individuals were based on allele frequencies for all individuals pooled. The distribution of estimates between individuals from the same location was similar to the distribution of estimates between individuals from different locations (Fig. S3A), but the within-location distribution had a longer upper tail with some high values (Fig. S3B). There were five pairs with estimates of r above 1/8 (0.125), the degree of relatedness expected for first cousins, although their 95% confidence limits included zero (Table 2). Each of these pairs of apparently related individuals were collected from the same location in the same year and were at the same life stage. This could suggest that closely related individuals settled in cohorts at the same locations. Alternatively, our estimates of relatedness could have been upwardly biased by being based on allele frequencies for individuals pooled from different locations, sampling times and life stages. We investigated this potential bias by estimating r with allele frequencies from more restricted samples: the pair's source location, the pair's combination of location and sampling year, and the pair's combination of location, year and life stage. (Only one additional estimate was made for pairs from the JAC sample, which consisted entirely of juveniles collected in 2010.) As shown in Table 3, estimates of r decreased with each restriction of the samples used for allele frequency estimates. This effect is not confined to the upper tails of the distributions of r : the distribution of estimates of r based on allele frequencies at specific locations tended to be lower than estimates based on allele frequencies for all locations combined (Fig. S4).

Relatedness was estimated for all pairs within each life stage (megalopae, juveniles and adults) from the five intensively sampled Louisiana locations. Allele frequencies were estimated separately for each life stage with collection locations and years pooled. The distributions of estimates of coefficients of relatedness (r) were similar among life stages (Fig. S5A), and no estimates for pairs of megalopae were above 1/8 (Fig. S5B, Table 4). Thus, we found no evidence for reproductive sweepstakes events, such as full or half-siblings in the

same cohort of megalopae or year class of juveniles.

4 | DISCUSSION

4.1 | Geographic Variation

We detected very little geographic heterogeneity in the distribution of genetic variation among North American blue crabs despite extensive sampling and considerable statistical power. Our estimate of overall F_{ST} was 0.0002; without the sample from JAC the estimate would have been 0.0001 and not significantly above zero. In two previous studies, estimates of F_{ST} were on the order of 0.01, 50 times higher than ours. McMillen-Jackson and co-workers (1994) estimated *Nm* (population size times migration rate) from F_{ST} to be between 11.1 and 19.0, from which their unreported F_{ST} estimates can be inferred to have been between 0.013 and 0.022. From RAD sequence data, Plough (2017) estimated F_{ST} between a sample of blue crabs from the North American Atlantic and a sample from the Gulf of Mexico at 0.01. It is possible that this discrepancy in the magnitude of F_{ST} estimates is due to accidents of sampling. McMillen-Jackson and co-workers (1994) used only nine polymorphic allozymes, a small sample of loci. Furthermore, they estimated F_{ST} with the software BIOSYS-1 (Swofford & Selander, 1981), which doesn't use the now widely adopted Weir and Cockerham (1984) correction for sampling variance in allele frequency estimates, and therefore is expected to produce upwardly biased estimates of F_{ST} . Plough (2017) surveyed only two North American locations, one may have been unusually divergent like our sample from JAC. Differences in estimates of F_{ST} can also reflect differences in statistical methodology. A survey of microsatellite variation in North American blue crabs yielded average pairwise F_{ST} estimates that ranged between 0.003 and 0.008 depending on methodological details (Macedo et al., 2019). None of the pairwise F_{ST} estimates in this study were significantly different from zero when the full set of putatively neutral loci was analyzed by the permutation test in Arlequin (Schneider et al., 2000).

Our study encountered several potential sources of error that could lead to the false appearance of genetic structure in blue crabs or other species. The first is misidentification of specimens. Megalopae of several congeners of *C. sapidus* are found in the northern Gulf of Mexico and are difficult to distinguish by morphology. This could explain some of the anomalous findings of an allozyme survey of blue crabs from the coast of Texas, which included large temporal differences in allele frequencies among cohorts of settling blue crab megalopae over a span of several months (Kordos & Burton, 1993). The observed pattern closely matches seasonal changes in abundance of the megalopae of a related species, *Callinectes similis* , and the megalopae of these species are difficult to distinguish (Sullivan & Neigel, 2017). Sullivan and Neigel (2017) also identified the megalopae of *C. rathbunae* and *C. danae* co-settling with *C. sapidus* . In one of our samples, 6 out of 41 megalopae that we had identified as *C. sapidus* by morphology matched the distinctive pattern of homozygous and unscorable loci seen in a specimen of *C. rathbunae* . Had we not removed these megalopae from our analysis we would have concluded that the sample was genetically differentiated from others and included closely related individuals. Our laboratory's previous survey of sequence variation in protein-coding genes of blue crabs found nuclear gene haplotypes in blue crabs from Venezuela that were not found in samples from North America (Yednock & Neigel, 2014). These individuals from Venezuela appear to represent either a divergent population of *C. sapidus* or a different species. They were automatically excluded from the present study because of their unusually low heterozygosity (which also suggests a different species), but if we had included them our estimate of F_{ST} among locations would have been much higher. Another potential source of error is null alleles. Yednock and Neigel (2014) suspected null alleles at the *ant* locus in blue crabs because of departures from Hardy-Weinberg proportions, and we confirmed this in the present study. In their preliminary analysis, inclusion of data for this locus created the appearance of significant population differentiation not seen with other loci, even after a correction for null alleles was applied (Yednock and Neigel, unpublished). From the above considerations, it appears that known sources of error in genetic marker data tend to inflate estimates of F_{ST} and create false appearances of population structure. Large numbers of marker loci are often viewed as providing "high resolution" analysis of population structure (e.g.

Davey & Blaxter, 2010) because they increase the statistical power to detect genetic differentiation, but they may also increase the sensitivity of statistical tests to errors in data (Chapuis & Estoup, 2006).

4.2 | Accuracy of Illumina Genotyping

SNP genotypes determined by the Illumina Infinium assay for our study were highly reproducible and in aggregate, the data were consistent with the low population differentiation expected for a species that experiences very high gene flow. Differences between genotypes determined by the Infinium assay and those determined by sequencing are explained as null alleles in the sequencing data. Individuals with anomalous genotypes can be readily explained as misidentified specimens. Although a direct comparison between the accuracy of our data and those based on other types of genetic markers in blue crabs is not feasible, our comparatively low estimates of F_{ST} imply greater accuracy. The biology of blue crabs suggestsF_{ST} should be very low, and errors in genetic marker data are expected to inflate estimatesF_{ST} . The distinction between no genetic differentiation and slight genetic differentiation is not trivial. As Palumbi (2003) pointed out in discussing the relevance of apparent "slight geographic differentiation" to the design of marine reserves "Interpreting the significance of this slight genetic signal has been difficult because even mild genetic structure implies very limited demographic exchange between populations, but slight differentiation could also be due to sampling error." We hope that our study will serve to stimulate interest in the use of Infinium genotyping for population genetic surveys, and that the error assessment methods we developed will prove useful.

4.3 | Temporal Variation without Sweepstakes Reproductive Success

Sweepstakes Reproductive Success (SRS) occurs when the progeny of a few lucky spawners comprise the majority of a cohort of settling individuals, drastically reducing effective population size and creating immediate genetic drift (Hedgecock & Pudovkin, 2011; Hedrick, 2005). SRS is often considered the de facto cause of temporal variation in allele frequencies in marine populations, although it can be difficult to rule out other mechanisms (Cornwell et al., 2016). We detected temporal variation in allele frequencies for blue crabs across years: among samples from all locations as well as among samples exclusively from Louisiana estuaries. We tested for two spatiotemporal patterns predicted to result from SRS: 1) genetic differentiation among cohorts (chaotic genetic patchiness), and 2) the occurrence of siblings or half-siblings that are the progeny of the lucky spawners (Hedgecock & Pudovkin, 2011; Selkoe, Gaines, Caselle, & Warner, 2006). We found no evidence of chaotic genetic patchiness among 21 collections of settling megalopae. However, we did not sample every settlement event so it is possible that we missed those coinciding with SRS. We therefore also tested for genetic differences among annual cohorts of juveniles. Juveniles represent all the successful settlement cohorts of the previous year, and so would include any from SRS events. However, because temporal variation in juveniles could also be caused by post-settlement mechanisms unrelated to SRS it is important to also test for the second prediction of SRS: the presence of full or half siblings within the samples exhibiting temporal variation. Although we did detect genetic heterogeneity among juveniles from different years, we did not find any instances of siblings or half-siblings among juveniles or indeed among any of the individuals that we genotyped. This suggests that a mechanism other than SRS caused temporal variation. We also uncovered a potential bias in relatedness estimation that can lead to the false appearance of closely related individuals. When we estimated coancestry with allele frequencies from pooled samples it appeared that pairs of closely related individuals occurred in some samples. However, this was an artifact created by differences in allele frequencies among samples, which causes individuals from the same sample to have more alleles in common than expected for unrelated individuals drawn from pooled samples. In this light, we consider the findings of Iacchei and co-workers (2013) who used microsatellite allele frequencies pooled from 17 locations to estimate kinship between lobsters (*Panulirus interruptus*). They reported average kinship varied among locations and that related individuals were more often at the same location than expected by chance. From this, they concluded that "siblings more likely settle together than disperse across sites". Our analysis suggests that observations like this could be artifacts.

5 | CONCLUSIONS

This study was unusual in combining the statistical power provided by over 1300 loci with the spatiotemporal coverage provided by sampling over 1400 individuals across space, time, and life stage. The finding of near genetic uniformity over large distances is an expected consequence of a life history that facilitates dispersal. Where genetic differentiation was detected among locations or among years, estimates of F_{ST} were on the order of 0.0001, below the detection limits of most previous studies. We did not determine the causes of these slight deviations from spatial and temporal homogeneity, but spatial variation was not related to distance and sweepstakes reproductive success does not appear to have been the cause of temporal variation. Gene flow from distant populations, selection, and conventional genetic drift remain as possible explanations. In a forthcoming paper, we will examine spatiotemporal patterns of variation at individual loci in relation to environmental variation and consider the possible role of natural selection.

ACKNOWLEDGEMENTS

This research was made possible by grants from the Gulf of Mexico Research Initiative (GRI-013 and GoMRI2012-II-523), the National Science Foundation (OCE-0315995 and NSF (2010)-PFUND-214), the Louisiana Board of Regents (ENH_2008–09) and by University of Louisiana at Lafayette Doctoral Fellowships to B.Y. and T.S. Special thanks to the Louisiana Department of Wildlife and Fisheries, the Army Corps of Engineers, the University of New Orleans, the Louisiana Universities Marine Consortium (LUMCON), the National Marine Fisheries Service, and the Texas Parks and Wildlife Department for assistance with sampling. We are grateful for sample collections provided by Mary Kay Fox and Chris Weidman at Waquoit Bay National Estuarine Research Reserve, Fred Scharf at University of North Carolina Wilmington, Kyle Spiller at Texas Parks and Wildlife Department, and Matt Watkins and Shannon Martin at Florida Fish and Wildlife Conservation Commission. Samples from the University of Louisiana at Lafayette Zoological Collection were kindly provided by Darryl Felder. We also thank Rachel Sullivan and Eric Dean for assistance with collecting. We thank Thomas Schultz Division at the Duke University Nicholas School of the Environment for making available to a draft genome sequence of the blue crab. We thank the DNA Technologies Core, and especially Siranoosh Ashtari, for help with the Infinium Assay genotyping.

REFERENCES

Aguilar, R., Hines, A. H., Wolcott, T. G., Wolcott, D. L., Kramer, M. A., & Lipcius, R. N. (2005). The timing and route of movement and migration of post-copulatory female blue crabs, Callinectes sapidus Rathbun, from the upper Chesapeake Bay. *Journal of Experimental Marine Biology and Ecology, 319* (1-2), 117-128.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215* , 403-410.

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

Berthelemy-Okazaki, N. J., & Okazaki, R. K. (1997). Population genetics of the blue crab *Callinectes sapidus* from the Northwestern Gulf of Mexico. *Gulf of Mexico Science, 1* , 35-39.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., . . . Nemesh, J. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genetics, 22* (3), 231.

Chapuis, M.-P., & Estoup, A. (2006). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution, 24* (3), 621-631.

Cornwell, B. H., Fisher, J. L., Morgan, S. G., & Neigel, J. E. (2016). Chaotic genetic patchiness without sweepstakes reproduction in the shore crab Hemigrapsus oregonensis. *Marine Ecology Progress Series, 548* , 139-152. doi:10.3354/meps11628

Davey, J. W., & Blaxter, M. L. (2010). RADSeq: Next-generation population genetics. *Briefings in Functional Genomics, 9* (5-6), 416-423.

Epifanio, C. E. (2007). Biology of Larvae. In V. S. Kennedy & L. E. Cronin (Eds.), *The Blue Crab: Callinectes sapidus* (pp. 513-533). College Park, Md.: Maryland Sea Grant College.

Feng, X., Williams, E. P., & Place, A. R. (2017). High genetic diversity and implications for determining population structure in the blue crab*Callinectes sapidus* . *Journal of Shellfish Research, 36* (1), 231-243.

Gelpi, C. G., Fry, B., Condrey, R. E., Fleeger, J. W., & Dubois, S. F. (2013). Using δ13C and δ15N to determine the migratory history of offshore Louisiana blue crab spawning stocks. *Marine Ecology Progress Series, 494* , 205-218.

Hedgecock, D., & Pudovkin, A. I. (2011). Sweepstakes reproductive success in highly fecund marine fish and shellfish: A review and commentary.*Bulletin of Marine Science, 87* (4), 971-1002.

Hedrick, P. (2005). Large variance in reproductive success and the Ne/N ratio.*Evolution, 59* (7), 1596-1599.

Hellberg, M. E. (2009). Gene Flow and Isolation among Populations of Marine Animals. *Annual Review of Ecology, Evolution, and Systematics, 40* , 291-310. doi:10.1146/annurev.ecolsys.110308.120223

Hellberg, M. E., Burton, R. S., Neigel, J. E., & Palumbi, S. R. (2002). Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science, 70* (1), 273-290 Suppl. S.

Iacchei, M., Ben-Horin, T., Selkoe, K. A., Bird, C. E., García-Rodriguez, F. J., & Toonen, R. J. (2013). Combined analyses of kinship and FST suggest potential drivers of chaotic genetic patchiness in high gene-flow populations. *Molecular Ecology, 22* (13), 3476-3494.

Johnson, M. S., & Black, R. (1982). Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. *Marine Biology, 70* (2), 157-164.

Johnson, M. S., & Black, R. (1984). Pattern beneath the chaos: The effect of recruitment on genetic patchiness in an intertidal limpet.*Evolution, 38* (6), 1371-1383.

Kennedy, V. S., & Cronin, L. E. (Eds.). (2007). *The Blue Crab: Callinectes sapidus* . College Park, Maryland: Maryland Sea Grant College.

Kingsford, M. J., Leis, J. M., Shanks, A., Lindeman, K. C., Morgan, S. G., & Pineda, J. (2002). Sensory environments, larval abilities and local self-recruitment. *Bulletin of Marine Science, 70* (1), 309-340.

Kordos, L. M., & Burton, R. S. (1993). Genetic differentiation of Texas Gulf coast populations of the blue crab *Callincectes sapidus* .*Marine Biology, 117* , 227-233.

Lacerda, A. L. F., Kersanach, R., Cortinhas, M. C. S., Prata, P. F. S., Dumont, L. F. C., Proietti, M. C., . . . D'Incao, F. (2016). High connectivity among blue crab (*Callinectes sapidus*) populations in the Western South Atlantic. *Plos One, 11* (4), e0153124.

Macedo, D., Caballero, I., Mateos, M., Leblois, R., McCay, S., & Hurtado, L. A. (2019). Population genetics and historical demographic inferences of the blue crab *Callinectes sapidus* in the US based on microsatellites.*PeerJ, 7* , e7780.

McMillen-Jackson, A. L., & Bert, T. M. (2004). Mitochondrial DNA variation and population genetic structure of the blue crab *Callinectes sapidus* in the eastern United States. *Marine Biology, 145* (4), 769-777.

McMillen-Jackson, A. L., Bert, T. M., & Steele, P. (1994). Population genetics of the blue crab *Callinectes sapidus* : Modest population structuring in a background of high gene flow. *Marine Biology, 118* (1), 53-65.

Metcalf, K. S., van Montfrans, J., Lipcius, R. N., & Orth, R. J. (1995). Settlement indices for blue crab megalopae in the York River, Virginia: Temporal relationships and statistical efficiency. *Bulletin of Marine Science, 57* , 781-792.

Milligan, B. G. (2003). Maximum-likelihood estimation of relatedness.*Genetics, 163* (3), 1153-1167.

[dataset] Neigel, J. E. (2017). Single nucleotide polymorphism genotypes of blue crabs, Gulf of Mexico The Environmental Effects of an Oil Spill on Blue Crabs in the Gulf of Mexico and the Dynamics of Recovery: Integrating Oceanography and Molecular Ecology. GRIIDC. https://data.gulfresearchinitiative.org. UDI: R2.x214.000:0008. DOI: 10.7266/N7JW8C72.

Newcombe, C. L., Sandoz, M. D., & Rogers-Talbert, R. (1949). Differential growth and moulting characteristics of the blue crab, Callinectes sapidus Rathbun.*Journal of Experimental Zoology, 110* (1), 113-152.

Palumbi, S. (2003). Population genetics, demographic connectivity and the design of marine reserves. *Ecological Applications, 13* (1), S146-S158.

Plough, L. V. (2017). Population genomic analysis of the blue crab*Callinectes sapidus* using genotyping-by-sequencing. *Journal of Shellfish Research, 36* (1), 249-262.

Rodrigues, M. A., Dumont, L. F. C., dos Santos, C. R. M., D'Incao, F., Weiss, S., & Froufe, E. (2017). Two distinct mtDNA lineages of the blue crab reveal large-scale population structure in its native Atlantic

distribution. *Estuarine, Coastal and Shelf Science, 197* , 45-53.

Rousset, F. (2008). GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources, 8* (1), 103-106.

Ryman, N. (2006). CHIFISH: A computer program testing for genetic heterogeneity at multiple loci using chi-square and Fisher's exact test.*Molecular Ecology Notes, 6* (1), 285-287. doi:10.1111/j.1471-8286.2005.01146.x

Ryman, N., & Palm, S. (2006). POWSIM: A computer program for assessing statistical power when testing for genetic differentiation.*Molecular Ecology Notes, 6* (3), 600-602. doi:10.1111/j.1471-8286.2006.01378.x

Ryman, N., Palm, S., Andre, C., Carvalho, G. R., Dahlgren, T. G., Jorde, P. E., . . . Ruzzante, D. E. (2006). Power for detecting genetic divergence: differences between statistical methods and marker loci. *Molecular Ecology, 15* (8), 2031-2045. doi:10.1111/j.1365-294X.2006.02839.x

Santos, C. R. M. d., & D'Incao, F. (2004). Crustaceos no cerrito Ariano Souza, Rio Grande, Rio Grande do Sul e distribuicao de *Callinectes sapidus* (Brachyura, Portunidae). *Iheringia. Serie Zoologia, 94* (1), 73-76.

Schneider-Broussard, R., Felder, D. L., Chlan, C. A., & Neigel, J. E. (1998). Tests of phylogeographic models with nuclear and mitochondrial DNA sequence variation in the stone crabs, *Menippe adina* and *Menippe mercenaria* . *Evolution, 52* (6), 1671-1678.

Schneider, S., Roessli, D., & Excoffier, L. (2000). Arlequin: A software for population genetics data analysis. Ver 2.000. (Version 2.000). Geneva, Switzerland: Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Selkoe, K. A., Gaines, S. D., Caselle, J. E., & Warner, R. R. (2006). Current shifts and kin aggregation explain genetic patchiness in fish recruits.*Ecology, 87* (12), 3082-3094.

Shaklee, J. B., & Bentzen, P. (1998). Genetic identification of stocks of marine fish and shellfish. *Bulletin of Marine Science, 62* (2), 589-621.

Sinnock, P. (1975). The Wahlund effect for the two-locus model. *The American Naturalist, 109* (969), 565-570.

Steemers, F. J., Chang, W., Lee, G., Barker, D. L., Shen, R., & Gunderson, K. L. (2006). Whole-genome genotyping with the single-base extension assay.*Nature Methods, 3* (1), 31.

Stephens, M., & Donnelly, P. (2003). A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics, 73* (5), 1162-1169.

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics, 68* (4), 978-989.

Sullivan, T. J., & Neigel, J. E. (2017). Misidentification of megalopae as a potential source of error in studies of population genetics and ecology of the blue crab Callinectes sapidus. *Marine Ecology Progress Series, 565* , 95-111. doi:10.3354/meps12043

Swofford, D. L., & Selander, R. B. (1981). BIOSYS-1: A FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. *Journal of Heredity, 72* (4), 281-283.

Van Engel, W. (1990). Development of the reproductively functional form in the male blue crab, Callinectes sapidus. *Bulletin of Marine Science, 46* (1), 13-22.

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

Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity, 89* (5), 438-450.

Waples, R. S. (2014). Testing for Hardy–Weinberg Proportions: Have We Lost the Plot? *Journal of Heredity, 106* (1), 1-19. doi:10.1093/jhered/esu062

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

Yednock, B. K., & Neigel, J. E. (2014). An investigation of genetic population structure in blue crabs, *Callinectes sapidus* , using nuclear gene sequences. *Marine Biology, 161* (4), 871-886. doi:10.1007/s00227-013-2387-1

Yednock, B. K., Sullivan, T. J., & Neigel, J. E. (2015). De novo assembly of a transcriptome from juvenile blue crabs (*Callinectes sapidus*) following exposure to surrogate Macondo crude oil. *BMC Genomics, 16* . doi:10.1186/s12864-015-1739-2

Data Accessibility

Data for this project is available from GRIIDC (https://data.gulfresearchinitiative.org), the database of the Gulf of Mexico Research Initiative as dataset UDI:R2.x214.000:0008.

Author Contributions

All authors contributed to the conception and design of the study. B.Y. and T.S. collected specimens. B.Y., T.S., and S.P. performed DNA extractions. S.P. designed the genotyping assays. V.R. helped perform the genotyping. S.P. and J.N. processed and analyzed data. J.N. and S.P. wrote the manuscript. All authors read and approved the final manuscript.

Tables

Table 1. Locations from which blue crabs (*Callinectes sapidus*) were collected and genotyped.

SITE N Place Name Country State Latitude Longitude

ATC 25 Atchafalaya USA LA 29.588 -91.228

CAL 26 Calcasieu Lake USA LA 29.838 -93.321

CDK 16 Cedar Key USA FL 29.093 -83.075

EMV 2 El Mojan Venezuela Zulia 11.339 -70.907

FWC 281 Freshwater City Locks USA LA 29.552 -92.306

GIL 253 Grand Isle LWDF Lab USA LA 29.239 -90.002

JAC 14 Jacksonville USA FL 30.408 -81.518

LSV 1 Laguna de Sontecompan, Los Tuxtlas Mexico Veracruz 18.550 -95.017

LUM 288 LUMCON Marine Laboratory USA LA 29.254 -90.664

MIL 44 Marsh Island USA LA 29.578 -91.884

PNT 18 Lake Pontchartrain USA LA 30.362 -90.166

RBC 144 Rutherford Beach Culvert USA LA 29.779 -93.133

RRC 16 Rhode River, Chesapeake Bay USA MD 38.892 -76.528

RWR 268 Rockefeller NWR USA LA 29.712 -92.766

ULM 16 Upper Laguna Madre USA TX 27.393 -97.223

WIL 16 Wilmington USA NC 33.961 -77.976

WQB 15 Waquoit Bay USA MA 41.563 -70.522

ZLV 3 Zamuro Venezuela Falcon 10.948 -71.193

Total 1446

Table 2. Pairs with estimated coefficient of relatedness (r) above 0.125.

Individual 1 Individual 2 estimated r 95% Confidence Interval

JAC-2010-Juv JAC-2010-Juv 0.1253 0 - 0.1559

JAC-2010-Juv JAC-2010-Juv 0.1304 0 - 0.1562

LUM-2010-Juv LUM-2010-Juv 0.1472 0 - 0.2050

LUM-2011-Adult LUM-2011-Adult 0.1258 0 - 0.1923

RBC-2014-Adult RBC-2014-Adult 0.2001 0 - 0.2722

Table 3. Estimates of relatedness (r) above 0.125 when based on overall allele frequencies and revised estimated based on allele frequencies estimated within locations, and for combinations of location and year, and for combinations of location, year and life-stage.

Allele Frequencies used for Estimates

Individual 1 Individual 2 Overall Location Location-Year Location-Year-Stage

JAC-2010-Juv JAC-2010-Juv 0.1253 NA NA 0.0000

JAC-2010-Juv JAC-2010-Juv 0.1304 NA NA 0.0123

LUM-2010-Juv LUM-2010-Juv 0.1472 0.142 0.1307 0.1092

LUM-2011-Adult LUM-2011-Adult 0.1258 0.1247 0.0885 0.0239

RBC-2014-Adult RBC-2014-Adult 0.2001 0.1882 0.1409 0.0377

Table 4. Mean and maximum estimates of relatedness (r) among all pairs of individuals at each live stage and the number pairs with r above 0.125. Allele frequencies were estimated for individuals pooled across locations and years.

Life Stage N Pairs Mean r Max r N above 0.125

Megalopae 54615 0.009392 0.1179 0

Juveniles 107416 0.010008 0.1438 1

Adults 93096 0.009638 0.1986 1

Figures

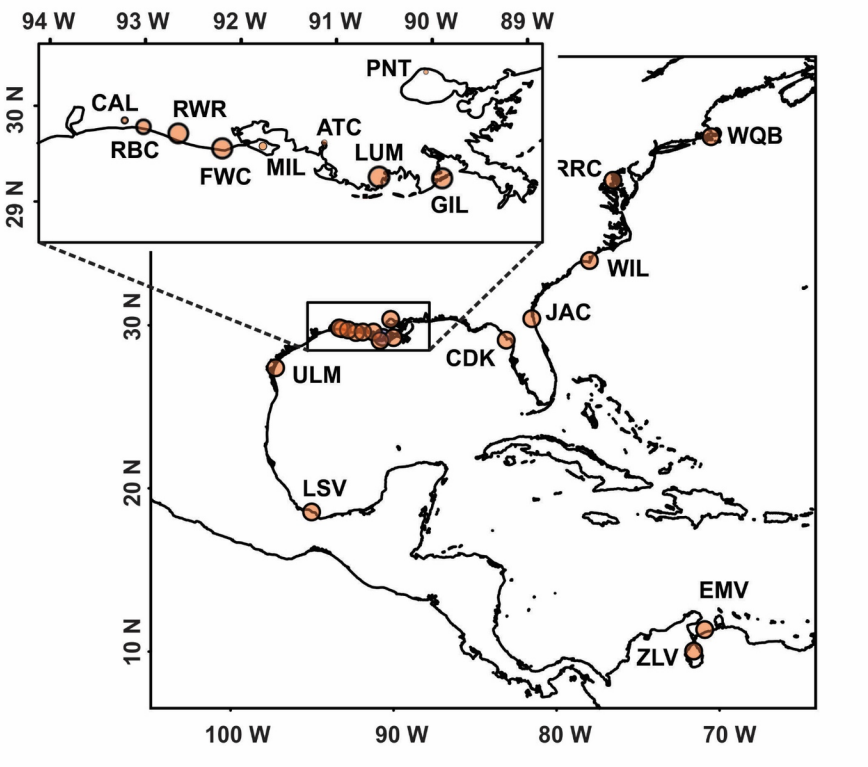

Figure 1. Sites from which blue crabs were collected. Inset: Sites along the Louisiana coast with the area of each circle proportional to sample size.

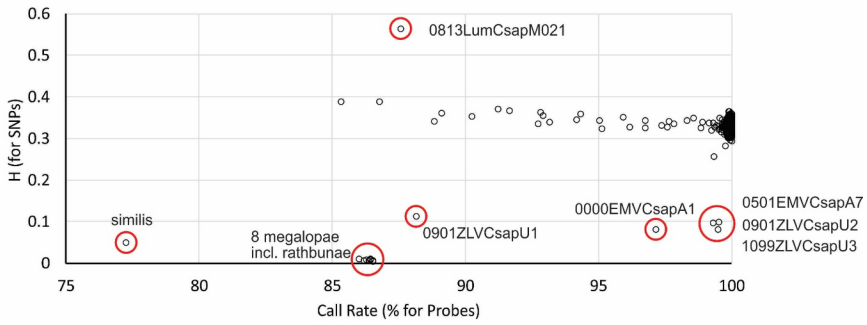

Fig. 2. Observed heterozygosity (for SNP loci) and call rates (for individual Infinium probes) for all individuals, including those assigned to species of *Callinectes* other than *C. sapidus* .