



OPEN Evidence of positive selection in the *COI* mitochondrial gene in the blue crab *Callinectes sapidus* (Decapoda: Portunidae)

Marco Mancuso¹, Francesco Tiralongo^{1,2,3}, Giada Santa Calogero¹, Gea Manganaro¹, Alessandro Nota^{2,4}, Venera Ferrito^{1✉} & Anna Maria Pappalardo^{1✉}

The Blue Crab (*Callinectes sapidus* Rathbun, 1896) is a decapod crustacean native to the Western Atlantic coast that has become a highly invasive presence in Mediterranean waters since the 1950s. This species incarnates multiple characteristics that render it a successful invader: large size, high reproductive output, a generalist diet, euryhalinity, and eurythermal physiology. This study aims to investigate the evolutionary implications of the colonization of Mediterranean ecosystems by looking at the evolution of the mitochondrial gene *COI* in invasive *C. sapidus*. After assessing phylogenetic relationships of native and invasive populations, signs of selection analyses were carried out, revealing a multitude of sites under positive selection in invasive populations. Our results might point toward an ongoing trend of adaptive evolution triggered by the selective pressure that a new environment exerts upon this invasive species in Mediterranean ecosystems.

Keywords Blue crab, *COI*, Positive selection, Invasive species

The Blue Crab (*Callinectes sapidus* Rathbun, 1896) is a commercially and ecologically relevant decapod crustacean belonging to the family Portunidae. *Callinectes sapidus* is a large crab species reaching 180 mm in carapace width in both sexes. Mature females carry a mass of eggs under their abdomens, containing between 700,000 and 2 million eggs¹. This species is euryhaline and thrives in both saltwater and freshwater ecosystems. *C. sapidus* is also highly eurythermic, being able to comfortably live in waters harboring temperatures ranging from 7 °C to 32 °C^{2,3}. The ability to adapt to different environments, paired with its size, generalist feeding behavior, and large reproductive potential, make the blue crab the real-life representation of the perfect invader³. While this species is originally native to the coastal waters of the western Atlantic Ocean from North to South America, it has established itself as an unwanted presence in Mediterranean waters as well. Its introduction dates back to the early 1900s, likely through ballast water transport and other anthropogenic activities. Ever since, this species has significantly altered native biodiversity, fisheries, and coastal ecosystems of its invaded range, constantly gaining ground over native species and disrupting local food webs in a wide array of aquatic ecosystems^{2,4,5}.

Given its significant ecological and economic consequences, genetic research on *C. sapidus* in the Mediterranean Sea is a rapidly growing field. Currently, available data on invasive blue crabs in Mediterranean waters belongs to specimens sampled in Albania, Greece, Italy, Spain, Turkey and Tunisia^{2,3,6–10}. Preliminary studies have suggested that Mediterranean populations may exhibit founder effects, characterized by reduced genetic diversity following the initial establishment of the species. It is crucial to investigate whether these populations show signs of genetic differentiation compared to their native Atlantic counterparts and whether they have undergone adaptation to the novel Mediterranean environment. Mitochondrial DNA can be of extreme interest for this purpose. Mitochondria are highly important cellular organelles implicated in multiple energy-related processes, such as the production of ATP, and several other biosynthetic pathways¹¹. Mitochondria are of symbiotic origin and retain their own DNA (~ 16–17 kb) that is exclusively maternally inherited, and presents no recombination; therefore, mutations are accumulated without recombination-level mechanisms correcting for deleterious ones^{12,13}. Due to all these characteristics, the mitochondrial genome is a useful tool

¹Department of Biological, Geological, and Environmental Sciences, University of Catania, Catania, Italy. ²Ente Fauna Marina Mediterranea, Scientific Organization for Research and Conservation of Marine Biodiversity, Avola, Italy. ³National Research Council, Institute of Marine Biological Resources and Biotechnologies, Ancona, Italy. ⁴Department of Veterinary Sciences, University of Turin, Grugliasco, Italy. ✉email: vferrito@unict.it; pappalam@unict.it

to investigate population genetics and adaptive evolution patterns. Of all mitochondrial genes, Cytochrome Oxidase I (*COI*) is surely one of the most extensively studied. The *COI* gene has been a staple component of most studies involving molecular phylogenetics, phylogeography, DNA barcoding, and speciation patterns in vertebrates and invertebrates for the past 30 years due to its relatively high mutation rate and omnipresence across organisms¹⁴. This gene is involved in cellular respiration and is part of Complex IV of the electron transport chain in mitochondria^{15,16}. Cytochrome C Oxidase I (*ccox1*), the protein encoded by *COI*, was found to be quite an important protein for marine estuarine species, often subject to hypoxic conditions, in which the protein is downregulated to cope with the lower levels of oxygen¹⁷. Therefore, it is possible that *C. sapidus*, an estuarine species itself, might have experienced some compositional and/or structural changes in *ccox1* to deal with anoxic conditions. This study aims to fill the abovementioned gaps in knowledge by exploring signs of selection in Mediterranean populations of *C. sapidus* using the mitochondrial DNA marker *COI*. By comparing these populations to those in its native Atlantic range, we aim to unravel the invasion history, determine the role of founder effects, and assess the potential action of selection pressure on mitochondrial OXPHOS genes.

Results

Haplotype designation and molecular phylogenetics

Our analysis in RStudio revealed a total of 198 *COI* haplotypes out of the 500 sequences we featured as input data (Supplementary Table 1). Private *COI* haplotypes (hereafter indicated in round brackets) were found in blue crab populations from Brazil (5), Costa Rica (1), Greece (2), Mexico (12), Nicaragua (1), Turkey (10), the USA (103), Venezuela (21), and, ultimately, Sicily (28), the focal sampling area for the present study. Additionally, 15 haplotypes were found to be shared among geographically segregated samples.

We found invasive populations from the Mediterranean Sea forming distinct clusters. Both of our phylogenies (Figs. 1 and 2) agree on the presence of two separate clusters featuring sequences from invaded zones across the Mediterranean Sea. The clusters we found mainly comprise haplotypes from Sicily, our main sampling area, with Greek and shared haplotypes from the Mediterranean invaded range (Spain, again Greece, and Turkey, as well as other Italian populations and the native USA range) fitting in between. However, a single haplotype (H26, from Sicily) deviated from this paradigm and was found nested within American samples (Figs. 1 and 2). Within haplotypes from the invaded Mediterranean range, the only ones not clustering with the rest are those from the Turkish Mediterranean coast, which show more affinity with the native American haplotypes.

Another evident cluster we noticed features haplotypes from the Central and South American countries of Costa Rica, Mexico, Brazil, Nicaragua, and Venezuela, in addition to a single sample from the USA and some shared haplotypes that feature sequences from Jamaica and the USA as well. American (USA) haplotypes then dominate the rest of the most basal portions of the trees obtained through both phylogenetic methods, with some Mexican haplotypes fitting in between (Figs. 1 and 2).

Bayesian phylogeny showing the relationships among *COI* haplotypes of *C. sapidus*. The color legend on the left, the same as that used in Fig. 1, shows the sampling locations of the haplotypes featured in the analyses. Support values at nodes are represented by posterior probabilities (only shown when above 90%), highlighted in bold text. In light blue, clusters of invasive Mediterranean haplotypes. The figure was edited following the pipeline mentioned in the “Bioinformatics” methodological section.

Signals of selection and 3D protein homology modelling

After ensuring no recombination was found in our dataset by using the GARD tool available in HyPhy, we tested our dataset for pervasive and episodic selection in PAML (Codeml) and HyPhy (FUBAR, MEME). Our one-ratio model analysis in Codeml revealed an overall $\omega = 0.05$ across our dataset, indicating strong conservative evolution (Table 1). Conversely, our branch model, in which the haplotypes from the Mediterranean populations were tested for adaptive evolution against the native ones, indicated a higher selection coefficient characterizing the invasive foreground lineages, with ω standing at 0.2. While this still indicates purifying selection, it significantly deviates from the result of the one-ratio model ($\omega_{\text{branch model}} = 0.2$ vs. $\omega_{\text{one-ratio model}} = 0.05$; Table 1).

Concerning site selection, the M8 vs. M7 analysis in Codeml highlighted no support for sites being under positive selection across the entire dataset. The computed LRT stood at 0.00373, which is far from the χ^2 critical value at $p < 0.05$ at degrees of freedom = 2 (c.v. = 5.991). Therefore, the null hypothesis (M7) could not be refuted, and, thus, positive selection at the site level cannot be considered supported. No site was, therefore, found under positive selection by Codeml under the BEB criterion in the M8 vs. M7 analysis on the full dataset. The HyPhy output does not deviate much from that of Codeml. Little evidence of positive selection was found at the site level. FUBAR indicated no evidence of pervasive selection for any site at posterior probability > 95%. On the other hand, 50 sites were signalled under negative selection, providing further evidence for a conservative evolutionary trend for the *C. sapidus* *COI* already suggested by Codeml. However, our analyses in MEME detected sites 110 and 195 as being under episodic positive selection ($p < 0.05$; Table 1), which means that, despite the overall trend depicting negative selection, instances of sites under episodic diversification are present.

Yet, the most interesting findings came from the branch-site model analyses done in Codeml. Here, we wanted to test adaptive evolution at the site level in branches representing invasive *C. sapidus* in Mediterranean waters. Our results indicated that 23.75% of sites were found to evolve under varying degrees of positive selection in foreground lineages exclusively. While this result sounds astounding, it is worth mentioning that, while computing this statistic, Codeml also features sites with PP < 95% within this category, inevitably inflating the results on the percentage of positively selected sites in foreground lineages. Out of these, we only considered sites evolving under strong positive selection (PP > 95%) and managed to find seven sites evolving under a strong diversifying framework in foreground lineages: 4, 75, 110, 148, 178, 195, and 199 (Table 1). However, as the Likelihood Ratio Test between the alternative and the null model revealed a non-significant result, therefore, we treat these 7 sites retrieved under positive selection as likely being the result of episodic selection.

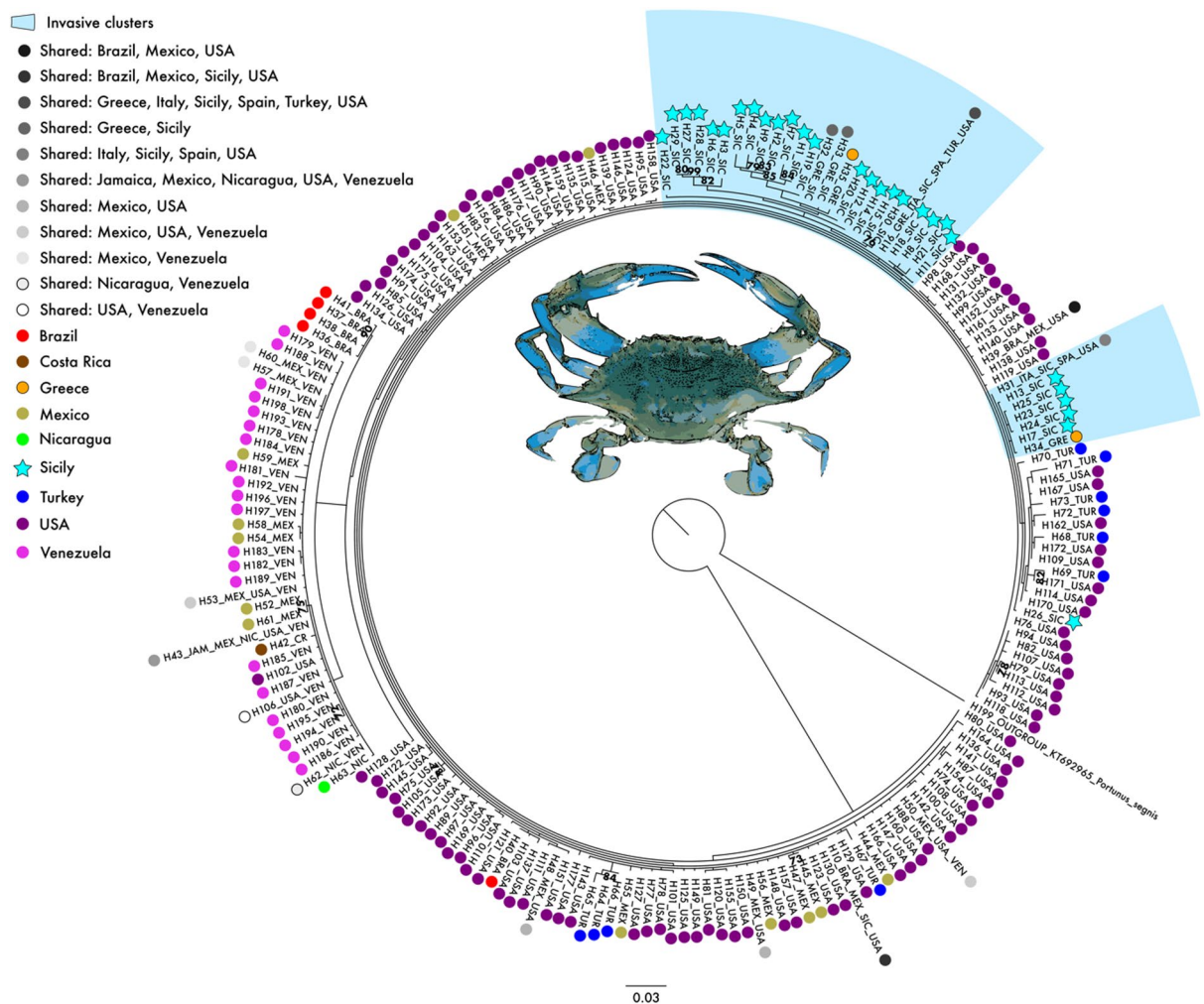


Fig. 1. Maximum Likelihood (ML) phylogeny of *COI* haplotype sequences of *Callinectes sapidus*. Maximum Likelihood phylogeny showing the relationships among *COI* haplotypes of *C. sapidus*. The colour legend on the left shows the sampling locations of each haplotype featured in the analyses. Support values at nodes are represented by bootstrap values (>70), highlighted in bold text. In light blue, clusters of invasive Mediterranean haplotypes. The figure was edited following the pipeline mentioned in the “Bioinformatics” methodological section.

Besides instances of positive selection, we also looked at non-synonymous amino acid changes to see if any site change was unique to any single population, but none were found.

In order to understand where the sites under positive selection were located, we generated three-dimensional protein models for *COI*. Sites under diversifying selection retrieved by MEME and Branch-Site Model A in Codeml were mapped in hotpink and green, respectively (Fig. 3). Sites 110 and 195, retrieved under positive selection by both MEME and Codeml (only in invasive populations), are located halfway through the protein (110), and in proximity to the C-terminal end of the protein (195), respectively. Amino acids alanine (A) and aspartic acid (D) were found at site 110, while isoleucine (I) and phenylalanine (F) were found at position 195. Overall, the sites under positive selection in the invasive subset found by Codeml are well-distributed across the *COI* protein molecule, although multiple positively selected sites are detected at the C-terminal end of the protein (178, 195, 199).

Discussion

The results obtained from the *COI* sequence analysis of *C. sapidus* from the Mediterranean Sea indicate a geographic isolation of the invasive blue crabs from their native range, resulting in genotypic peculiarities that led to a clear separation of the Mediterranean haplotype clusters from the American ones under both ML and BI phylogenetic reconstructions. Another striking feature that suggests a high degree of diversification between native and invasive populations is the abundance of private haplotypes, which was especially noticeable across the range that we sampled. Among the 54 samples from Sicily (49) and Greece (5) that we analyzed for the present study, we identified 28 private haplotypes unique to Sicily, meaning that 57% of all the haplotypes from the newly sampled specimens from Sicily were exclusive to this region, and 2 that were unique to Greece, respectively.

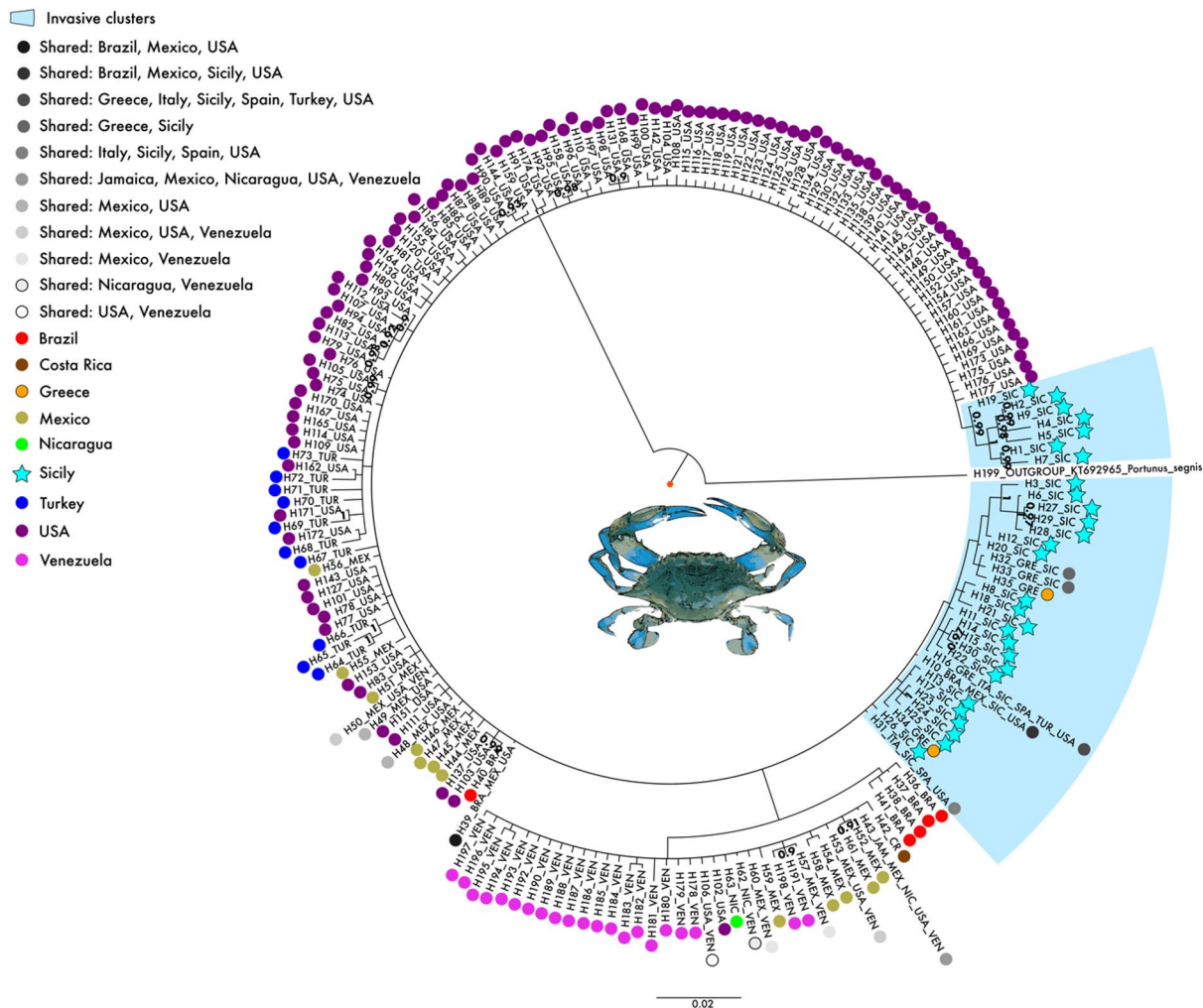


Fig. 2. Bayesian Inference (BI) phylogeny of COI haplotype sequences of *Callinectes sapidus*.

One-ratio model	Branch Model	Site models			Branch-Site model A
		Sites under positive selection			
		FUBAR (PP > 95%)	MEME (p < 0.05)	Codeml (PP > 95%)	Sites under positive selection (pp > 95%)
Overall ω	Foreground ω	–	110, 195	–	4, 75, 110, 148, 178, 195, 199

Table 1. Signs of selection analyses. PP = posterior probability; p = p-value significance. Branch-site model A output. Site classes: 0 = sites evolving under negative selection; 1 = sites evolving under a neutral evolutionary framework (dN/dS = 1); 2a = sites evolving under purifying selection in background lineages and under positive selection in foreground branches; 2b = sites evolving under neutral evolution in background lineages and under positive selection in foreground branches. The proportion of sites under positive selection under class 2a includes sites with PP < 95% as well.

Invasive populations of *C. sapidus* likely arrived in Mediterranean waters through ballast water from North American cargo ships^{2,18}. Multiple reports from the 20th century, dating back to 1901, suggest this species already roamed European waters off coastal France at the beginning of the twentieth century¹⁹. However, the first reports from Mediterranean waters date back to the 1950s^{2,20–22}. This species settled in its invaded range through several colonization events, which inevitably represented separate bottlenecks. A bottleneck is a genetic phenomenon in which a subset of a larger population is isolated, leading to a decrease in overall heterozygosity through the loss of several alleles and the retention of others, which will constitute the basis for future diversification of the isolated population²³. Bottlenecks can impact the rise, frequency, and fixation rates of mutations. Invasive populations may retain mutations that increase fitness in a new environment and experience a spike in genetic variance available for diversification as a result of multiple founder events that might increase genetic variability

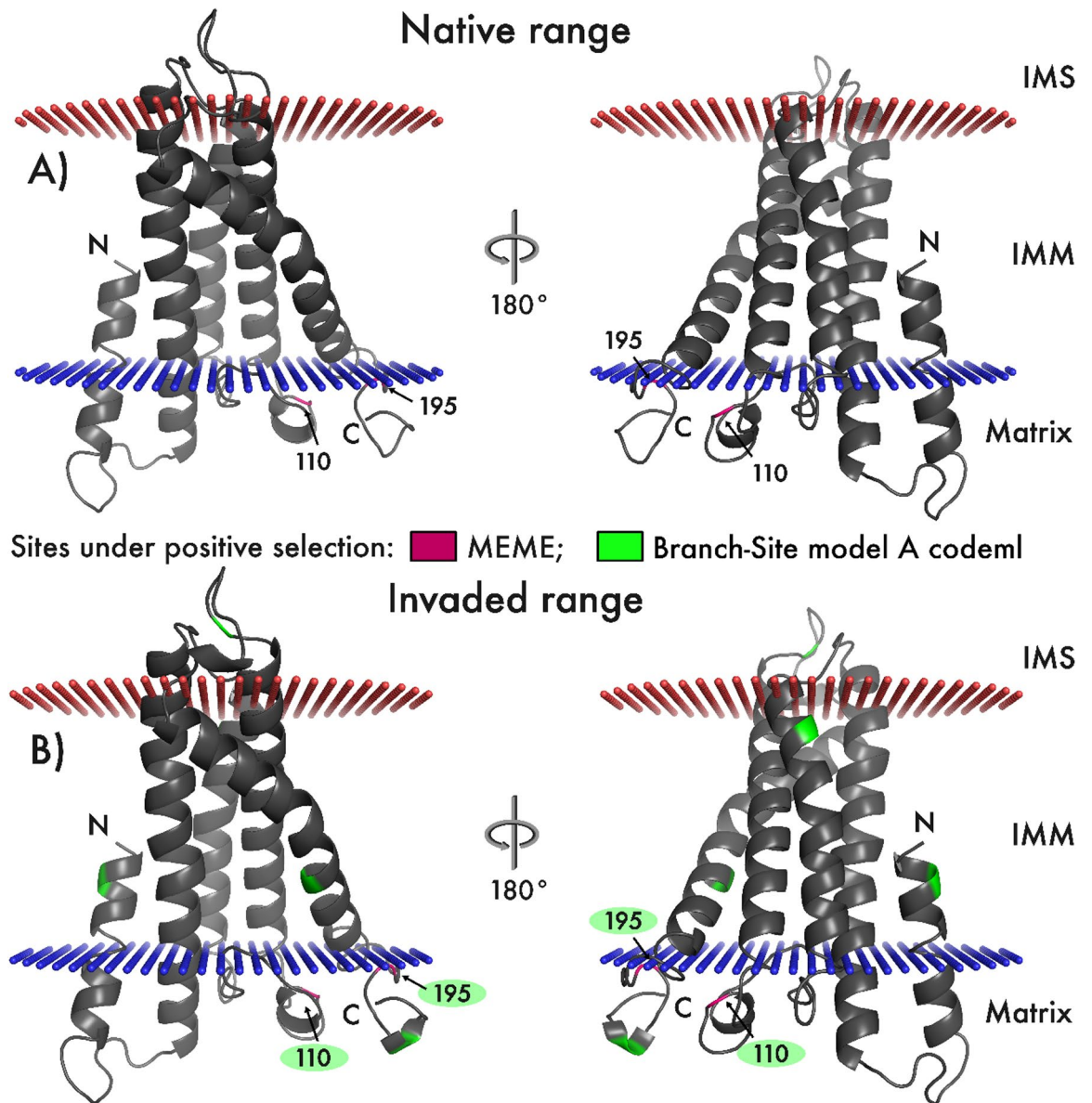


Fig. 3. Mapping of the positively selected sites in the three-dimensional model of the blue crab *Callinectes sapidus* COI protein from (A) native range and (B) invaded range. Sites identified by MEME are highlighted in purple; sites identified by the Codeml Branch-Site model A are highlighted in green. IMM = Inner mitochondrial membrane. IMS: intermembrane space. N, amino-terminal tail; C, carboxy-terminal tail. Codon positions 110 and 195, identified under positive selection by both MEME and branch-site model A in Codeml, are pointed at by black arrows. Green ellipses were placed around sites 110 and 195 from the invaded range to indicate that both methodologies found those positions to be under positive selection.

in the invaded range^{24–26}. However, the genetic variability of invasive species could derive from the propagule pressure, as repeated introduction events from multiple source populations can greatly enhance the genetic pool of the established population. High propagule pressure not only increases the likelihood of survival and establishment by overcoming demographic and environmental stochasticity but also promotes the admixture of diverse genotypes. This process may generate novel genetic combinations that facilitate local adaptation and rapid evolutionary responses to the new environmental conditions. The results from Sicily are particularly noteworthy when compared to previous studies focusing on other invaded Mediterranean territories^{2,6}, which do not show nearly as many private haplotypes as Sicily does. Locci et al. (2024) featured a total of 83 invasive Blue Crabs samples from Greece (15), the Levantine Sea region of Turkey (36), Spain (5), Peninsular Italy (2), Sardinia (21), and Sicily (4). Their findings show a much lower percentage of private haplotypes in every single one of those regions with a significant number of samples, with Greece only showing one unique haplotype out of 15 samples (6%), Sardinia only comprising two private haplotypes out of 21 samples (9%), for example. Turkey, which featured 11 private haplotypes out of 36 samples (34%) as found in², is an exception likely resulting from longer

Sampling location	Coordinates	Samples
ORN Vendicari, IT	36°48'09.5"N 15°05'56.8"E	CAL1-7; CAL47; CAL51
ORN Foce Fiume Simeto, IT	37°23'56.9"N 15°04'20.7"E	CAL8
Fiume Ciane, IT	37°02'31.2"N 15°14'05.5"E	CAL9
Augusta, IT	37°14'42.7"N 15°13'36.6"E	CAL10-24; CAL26-38; CAL41-46; CAL52-56
Nea Kamarina, GR	38°55'16.2"N 20°50'24.3"E	CAL57-61

Table 2. Sampling information on the specimens of *Callinectes Sapidus* processed for the present study. ORN = Oriented Natural Reserve.

isolation times and commercial reintroductions for fishing purposes². Turkish haplotypes were also the only Mediterranean ones consistently not clustering with the rest of the invasive populations in both of our analyses, marking a clear difference between them, nested within American samples, and the rest, comprising distinct groups of their own. Gonzalez-Ortegón et al. (2022) sampled 149 specimens of *C. sapidus* at three locations in Spain, only retrieving two haplotypes (CSWM1 and CSWM2). These two haplotypes were found to be shared among crabs from multiple Mediterranean sampling sites featured in this study, including our *de novo* samples from Sicily, which cluster with CSWM1 and CSWM2 (haplotypes 16 and 31, respectively, in the present study; Supplementary Table 1). These findings suggest a peculiar evolutionary trajectory tending toward diversification that is particularly remarkable in the invaded range of Sicily.

One explanation might lie in the areas in which Blue Crabs were sampled for the present study. Sampling sites of Augusta and ORN Fiume Ciane (SE Sicily) lie within a 55 km radius from of the largest petrochemical plants in Europe, located in Priolo Gargallo. This area is considered a site of high environmental risk due to the pollutants produced by nearby industries, including large quantities of mercury (Hg) and organic compounds such as hexachlorobenzene (HCB), polycyclic aromatic hydrocarbons (PAH), and polychlorinated biphenyls (PCBs)^{27–29}. Yet, *C. sapidus* thrives in such a suboptimal habitat type, as can be attested by the large quantity of samples from the Augusta site especially, with 36 of our samples being native to this area.

Evidence of adaptive evolution might also be observed in the results of our analyses investigating signs of selection. When the entire dataset was tested altogether, no evidence of positive selection could be found (global $\omega = 0.05$). The evolutionary trend highlighted by the one-ratio model can be explained by (A) the low variation rate across our sequences, which belong to haplotypes from different populations of the same species; and (B) the physiological role of COI in the electron transport chain^{2,2}. As the COI protein is involved in a crucial and highly conserved metabolic pathway, it is unlikely for mutations to be retained in order not to interfere with natural organismal physiology, unless strong selective pressure for mutation is exerted. On the other hand, when the dataset was sub-partitioned into native and invasive *C. sapidus*, the selection coefficient associated with the foreground invasive branches stood at 0.2, which, while still indicating strong purifying selection, is four times larger than what is computed for the entire dataset. These results highlight instances of diversification and, possibly, adaptive evolution in the branches corresponding to the invasive population of *C. sapidus* sampled across Mediterranean territories. Evidence of adaptive evolution can also be found when looking at site selection: while only two sites were flagged under episodic positive selection from MEME (110; 195) across the entire dataset, Mediterranean invasive blue crabs were found to show seven sites likely under episodic positive selection (> 95% PP) when partitioned against native samples (Table 1).

The results from the branch-site model, paired with those of the branch model, might highlight adaptive evolution in *C. sapidus* individuals from the Mediterranean Sea. However, our results rely on a single 609 bp Folmer fragment of the mitochondrial COI gene, whose short length might reduce its power to distinguish true selection from stochastic noise. An alternative explanation is that the observed patterns of Mediterranean haplotype diversity reflect multiple founder events. In this case, the link between bottlenecks, multiple introductions and invasion success should be considered since multiple introductions could contribute to increase genetic diversity in introduced populations^{26,30}. Broader native-range sampling and multilocus data will allow to define whether the origin of this mitochondrial diversity lies in the native range of the species or invaded areas. Moreover, this will allow to better identify the source populations of Mediterranean blue crabs and identify the potential routes of introduction for the species.

Methods

Sampling, DNA extraction, and sequence amplification

Fifty-four individuals of *C. sapidus* (Decapoda: Portunidae) were sampled in the Ionian Coast of Sicily at different localities and labelled under the serial code “CAL” (Table 2) (CAL1-7; CAL47-51: ORN Vendicari (September 2023; May 2024); CAL8: ORN Foce Fiume Simeto (October 2023); CAL9: ORN Fiume Ciane (November 2023); CAL10-24; CAL26-38; CAL41-46; and CAL52-56: Augusta (November 2023 - May 2024); CAL57-61: Nea Kamarina, Greece (September 2023)). Prior to laboratory processing, muscle samples from each specimen were placed in 96% ethanol. Samples were processed strictly following the guidelines set by the scientific ethical committee of the University of Catania (Italy). Genomic DNA extraction was performed using the NucleoSpin Tissue extraction kit (MACHEREY-NAGEL, Duren, Germany) following the manufacturer’s instructions starting from 20 to 30 mg of crab muscle tissue. After extraction, DNA concentration and purity (OD260/230 and OD260/280) were assessed through a spectrophotometric assay using Nanodrop-ONE (ThermoFisher, Wilmington, DE, USA). Mitochondrial DNA sequences corresponding to COI gene were isolated through Polymerase Chain Reaction (PCR). PCRs were performed using the Platinum Taq DNA Polymerase (Invitrogen,

Waltham, MS, USA) and carried in 25 μ L volumes. For primers and PCR thermal cycles see Vitale et al. 2015³¹. DNA concentration was kept between 100 and 400 ng/ μ L across the reactions to ensure successful amplification. A 1% electrophoresis gel was subsequently used to ensure correct amplification. PCR products and forward and reverse primers were outsourced at MacroGen Europe (Milan Genome Centre, IT) for Sanger sequencing.

Sequence gathering and preparation

Complementary *C. sapidus* COI sequences were gathered from NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>; accessed on multiple dates) to be compared with our *de novo* sequence data. A total of 500 *C. sapidus* COI sequences were gathered from NCBI GenBank, representing specimens sampled in Brazil, Costa Rica, Greece, Italy, Jamaica, Mexico, Nicaragua, Spain, Turkey, USA, and Venezuela. It should be noted that an important number of short COI sequences (around 436–574 bp) obtained from *C. sapidus megalops* from the Gulf of Mexico were deposited in Genbank by Grey and Franco (unpublished). However, we decided against using these sequences because they were too short and this would have resulted in an even shorter length of sequences in our dataset. One sequence (GenBank accession code: KT692965) of *Portunus segnis* (Decapoda: Portunidae) was downloaded as well to provide an outgroup for proper tree rooting. All sequences are available in Supplementary Table 1. Once total data was gathered, our *de novo* sequences of 658 bp were aligned to those downloaded in NCBI GenBank in MAFFT v.7.490³². The multialignment was downloaded from MAFFT in FASTA format and trimmed to 609 base pairs in AliView v.1.43³³. The number of haplotypes was then obtained in RStudio (version 2024.09.1 + 394)³⁴ using the package haplotypes³⁵.

Bioinformatics: molecular phylogenetics and inference of signs of selection

Phylogenetic relationships among *C. sapidus* haplotypes were inferred under both Bayesian Inference (BI) and Maximum Likelihood (ML) frameworks. The model of sequence evolution was inferred in JModelTest2³⁶ under default parameters. Both the corrected Akaike Information Criterion (AICc) and Bayesian Information Criterion (BIC) suggested a two-state substitution model with a gamma distribution (HKY + G) as the most appropriate fit for the data. The outgroup species *P. segnis* (KT692965) was included to root the tree. The BI analysis was performed in MrBayes v. 3.27³⁷ under the following Markov Chain Monte Carlo (MCMC) specifics: four independent swapping chains (nrns = 4, nchains = 4) ran until convergence was reached, for a total of 60,228,000 generations. Sampling was performed every 100 generations, with diagnostic output recorded at intervals of 1000 generations. The analysis was set to stop once the average standard deviation of split frequencies (ASDSF) had gone below a 0.01 threshold, which unequivocally suggests convergence. As an additional measure of convergence, we looked at the potential scale reduction factor (PSRF) parameter, which indicates convergence when approaching 1, which is what we found in our analysis. The burn-in was set to 25% (burninfrac = 0.25), and the first 25% of the samples were discarded to ensure the sole retention of the most reliable post-burn-in samples. All results were summarized under the 50% majority consensus rule.

Maximum Likelihood phylogenetic inference was performed in Raxml-ng v.1.2.2³⁸. Tree search began with 25 parsimony-based trees and 25 random trees. A total of 10,000 iterations were implemented for this analysis, and bootstrapping was enforced as a measure of node robustness.

Phylogenies were then visualized and edited in Figtree v.1.4.5³⁹, and subsequently arranged in publication-ready panels in Inkscape v.1.4⁴⁰.

Signs of selection analyses—site models

Signs of selection are a measure of how changes at the codon level might reflect evolutionary patterns. The measure of natural selection is a coefficient (ω) resulting from the ratio of non-synonymous (dN) mutations over synonymous (dS) mutations. Non-synonymous mutations imply a change at any of the three codon positions resulting in a different translated amino acid. On the other hand, synonymous mutations describe a change at any codon position resulting in no change in the translated amino acid. Therefore, a higher amount of non-synonymous mutations is a symptom of diversifying evolution (positive selection), resulting in $\omega > 1$. Conversely, if synonymous mutation prevails, the selection coefficient ω will be below 1, indicating a conservative evolutionary trend (negative selection). Finally, if dN equals dS, a neutral evolution scenario is observed. Natural selection in the COI protein-coding gene was investigated in HyPhy via the tools FUBAR and MEME⁴¹, as well as in Codeml, an available plug-in from PAML (v.4.10.7)⁴². Prior to submitting our analyses for signs of selection analyses, the outgroup sequence from *P. segnis* (GenBank accession code: KT692965) was pruned out of the input sequence data and the input Maximum Likelihood phylogenetic tree so that our analyses could be solely focused on *C. sapidus* COI evolution.

Our dataset was investigated for instances of positive selection in Codeml, one of the plug-ins available in PAML. Codeml is a versatile tool able to investigate signs of selection under branch models, site models, and even branch-site models⁴². We first implemented a so-called “one-ratio” model (specified via Model = 0, NSSites = 0 in the control file) to obtain a global selection coefficient calculated across the entire tree. Then, we tested for site selection globally using a combination of two alternative site models that are often tested together: M7 (beta) and M8 (beta & $\omega > 1$). These site models are tested against one another by specifying Model = 0 (an input that specifies the calculation of a single ω coefficient for the entire tree) and NSSites = 7 8 in the Codeml input file. M7 is treated as a null hypothesis model: it constrains ω between 0 and 1 at single sites, thus not supporting any hypothetical instance of positive selection. On the other hand, model 8 allows ω to exceed 1, and, therefore, admits the possibility of instances of positive selection. The two models are compared through a likelihood ratio test (LRT) at degrees of freedom = 2. The LRT result is then compared to the χ^2 critical value at $p < 0.05$ (critical value (c.v.) = 5.991), and, if the critical value is exceeded, then the null hypothesis of model 7 is rejected, providing support for positive selection at sites. Positively selected sites are identified via the Bayes Empirical Bayes (BEB) implemented in Codeml⁴³.

To further highlight possible instances of site selection, our dataset was analyzed in Hyphy to detect pervasive and episodic selection. Pervasive selection was investigated under a Bayesian framework using the Fast Unconstrained Bayesian AppRoximation FUBAR⁴⁴ package by specifying a 1000-point 50 × 50 grid to enhance precision. FUBAR computes a single omega coefficient for the entire input data, which enables the inference of both positively and negatively selected sites. A posterior probability of 95% was set as the threshold for defining a site as being under positive or negative selection.

Episodic selection was investigated under a maximum likelihood framework through the Mixed Effects of Model Evolution (MEME) package⁴¹. Unlike FUBAR, MEME calculates an ω coefficient for each branch. This method allows for detecting if a site is under positive selection even if it is not consistently selected positively over the entire tree. To increase robustness, we enforced 100 bootstrap resamples and specified corrections for double and triple hypothetical substitutions, allowing MEME to test for sites that might have encountered double or triple substitution events. Sites with a p-value below a 0.05 threshold were considered under positive selection.

Before performing all signs of selection analyses, the dataset was inspected in GARD (Genetic Algorithm for Recombination Detection)⁴⁵, to ensure no signs of recombination were found.

Signs of selection analyses—branch- and branch-site models

In addition to site selection at a global scale, we wanted to test whether the invasive populations in Mediterranean waters showed evidence of adaptive evolution. To do so, we implemented branch- and branch-site models available in Codeml, the same PAML plug-in we used to infer global site selection⁴². We looked for evidence of diversifying (= positive) selection by comparing the Mediterranean invasive populations (= foreground branches) to native lineages (= background branches) at both the branch and site levels. Before performing the analyses, we modified our ML input tree by partitioning it into two labelled subsets: foreground (label: “#1”; invasive blue crabs) and background (no label; native Western Atlantic blue crabs) lineages. Branch labelling was carried out at <https://phylotree.hyphy.org/>. The tree partitions we specified can be visualized in Supplementary Fig. 1. Branch models allow ω to vary among branches in the phylogeny. They are used to detect signs of selection in foreground branches, which are tested against a background branch set. To test for branch selection, we specified Model = 2 and NSSites = 0 in Codeml. The result is two omega values, one for the background and one for the foreground set. Branch-site models, on the other hand, allow ω to vary among both sites and tree branches. These models can be used to test site selection in a specific set of foreground lineages. We implemented branch-site model A (Model = 2; NSSites = 2) to be tested, allowing ω to be estimated freely across the foreground lineage. This test model is paired with a null model in which the selection coefficient omega is constrained (fix_omega = 1 and omega = 1 in Codeml), which allows for neutral evolution at most. The significance of the model is then assessed by performing a Likelihood Ratio Test, with results being labelled significant when exceeding the critical value 5.99 at $p < 0.01$. Again, sites under positive selection are inferred through the Bayes Empirical Bayes (BEB) algorithm available in Codeml⁴³.

Tridimensional protein modelling

Two ribbon structures of translated COI proteins were modelled in three dimensions in SwissModel⁴⁶. Sequences corresponding to Haplotype 47 (encompassing samples from Jamaica, Mexico, Nicaragua, USA, Venezuela) and Haplotype 12 (representing Sicilian sequences exclusively) were used as input, and then a *Callinectes ornatus* three-dimensional COI structure (accession code: A0A291NWD2_9EUC) previously modelled in AlphaFold⁴⁷ was used as a template to generate the 3D structures of both native and invasive *C. sapidus* *ccox1*. Spatial orientation of the protein was obtained by submitting the protein data bank (.pdb) file to the Orientation of Proteins in Membranes (OPM) database⁴⁸. The OPM output was then edited in PyMol⁴⁹.

Data availability

The COI sequence dataset generated in this study is deposited in GenBank at NCBI repository [Accession numbers: PV875401–PV875437]. Accession number of other sequences analyzed during the current study are available in Supplementary table I.

Received: 27 August 2025; Accepted: 27 November 2025

Published online: 07 December 2025

References

- Feng, X., Williams, E. P. & Place, A. R. High genetic diversity and implications for determining population structure in the blue crab *Callinectes Sapidus*. *J. Shellfish Res.* **36** (1), 231–242. <https://doi.org/10.2983/035.036.0126> (2017).
- Locci, C. et al. A sister species for the blue Crab, *Callinectes sapidus*? A Tale revealed by mitochondrial DNA. *Life* **14** (9), 1116. <https://doi.org/10.3390/life14091116> (2024).
- Schubart, C. D. et al. Phylogeography of the Atlantic blue crab *Callinectes Sapidus* (Brachyura: Portunidae) in the Americas versus the mediterranean sea: determining origins and genetic connectivity of a Large-Scale invasion. *Biology* **12** (1), 35. <https://doi.org/10.3390/biology12010035> (2022).
- Canning-Clode, J., Fowler, A. E., Byers, J. E., Carlton, J. T. & Ruiz, G. M. Caribbean creep’ chills out: climate change and marine invasive species. *PLoS ONE* **6** (12), e29657. <https://doi.org/10.1371/journal.pone.0029657> (2011).
- Tiralongo, F., Marcelli, M., Anselmi, G., Gattelli, R. & Felici, A. Invasion of freshwater systems by the Atlantic blue crab *Callinectes Sapidus* Rathbun, 1896 – new insights from Italian regions. *Acta Adriat.* **65** (2), 239–248. <https://doi.org/10.32582/aa.65.2.7> (2024).
- González-Ortegón, E. et al. Free pass through the pillars of hercules? Genetic and historical insights into the recent expansion of the Atlantic blue crab *Callinectes Sapidus* to the West and the East of the Strait of Gibraltar. *Front. Mar. Sci.* **9**, 918026. <https://doi.org/10.3389/fmars.2022.918026> (2022).
- Keskin, E. & Atar, H. H. DNA barcoding commercially important aquatic invertebrates of Turkey. *Mitochondrial DNA* **24** (4), 440–450. <https://doi.org/10.3109/19401736.2012.762576> (2013).

8. Vecchioni, L., Russotto, S., Arculeo, M. & Marrone, F. On the occurrence of the invasive Atlantic blue crab *Callinectes Sapidus* Rathbun 1896 (Decapoda: brachyura: Portunidae) in Sicilian inland waters. *Nat. History Sci.* <https://doi.org/10.4081/nhs.2022.586> (2022).
9. Vella, A. et al. New records of *Callinectes Sapidus* (Crustacea, Portunidae) from Malta and the San Leonardo river estuary in Sicily (Central Mediterranean). *Diversity* **15** (5), 679. <https://doi.org/10.3390/d15050679> (2023).
10. Besbes, N. et al. Molecular barcoding identification of the invasive blue crabs along Tunisian Coast. *Fishes* **9** (12), 485. <https://doi.org/10.3390/fishes9120485> (2024).
11. Kuznetsov, A. V. & Margreiter, R. Heterogeneity of mitochondria and mitochondrial function within cells as another level of mitochondrial complexity. *Int. J. Mol. Sci.* **10** (4), 1911–1929. <https://doi.org/10.3390/ijms10041911> (2009).
12. Cornuet, J. M. & Garnery, L. Mitochondrial DNA variability in honeybees and its phylogeographic implications. *Apidologie* **22** (6), 627–642. <https://doi.org/10.1051/apido:19910606> (1991).
13. Osellame, L. D., Blacker, T. S. & Duchon, M. R. Cellular and molecular mechanisms of mitochondrial function. *Best Pract. Res. Clin. Endocrinol. Metab.* **26** (6), 711–723. <https://doi.org/10.1016/j.beem.2012.05.003> (2012).
14. Pentinsaari, M., Salmela, H., Mutanen, M. & Roslin, T. Molecular evolution of a widely-adopted taxonomic marker (COI) across the animal tree of life. *Sci. Rep.* **6** (1), 35275. <https://doi.org/10.1038/srep35275> (2016).
15. Darling, J. A., Tsai, Y. H. E., Blakeslee, A. M. H. & Roman, J. Are genes faster than crabs? Mitochondrial introgression exceeds larval dispersal during population expansion of the invasive crab *Carcinus maenas*. *Royal Soc. Open. Sci.* **1** (2), 140202. <https://doi.org/10.1098/rsos.140202> (2014).
16. Willett, C. S. & Burton, R. S. Evolution of interacting proteins in the mitochondrial electron transport system in a marine copepod. *Mol. Biol. Evol.* **21** (3), 443–453. <https://doi.org/10.1093/molbev/msh031> (2004).
17. Brown-Peterson, N. et al. Molecular indicators of hypoxia in the blue crab *Callinectes Sapidus*. *Mar. Ecol. Prog. Ser.* **286**, 203–215. <https://doi.org/10.3354/meps286203> (2005).
18. Nehring, S. Invasion History and Success of the American Blue Crab *Callinectes sapidus* in European and Adjacent Waters. In B. S. Galil, P. F. Clark, & J. T. Carlton (Eds.), *In the Wrong Place—Alien Marine Crustaceans: Distribution, Biology and Impacts* (pp. 607–624). Springer Netherlands. (2011). https://doi.org/10.1007/978-94-007-0591-3_21
19. Bouvier, E. L. Sur Un *Callinectes Sapidus* M. Rathbun trouvé à Rochefort. *Bull. Mus. Hist. Nat. Paris.* **7**, 16–17 (1901).
20. Mancinelli, G., Bardelli, R. & Zenetos, A. A global occurrence database of the Atlantic blue crab *Callinectes Sapidus*. *Sci. Data.* **8** (1), 111. <https://doi.org/10.1038/s41597-021-00888-w> (2021).
21. Mancinelli, G. et al. The Atlantic blue crab *Callinectes Sapidus* in Southern European coastal waters: Distribution, impact and prospective invasion management strategies. *Mar. Pollut. Bull.* **119** (1), 5–11. <https://doi.org/10.1016/j.marpolbul.2017.02.050> (2017).
22. Ribeiro, F. & Verissimo, A. A new record of *Callinectes Sapidus* in a Western European estuary (Portuguese coast). *Mar. Biodivers. Records.* **7**, e36. <https://doi.org/10.1017/S1755267214000384> (2014).
23. Nei, M., Maruyama, T. & Chakraborty, R. The bottleneck effect and genetic variability in populations. *Evolution* **29** (1), 1. <https://doi.org/10.2307/2407137> (1975).
24. Carson, H. L. Increased genetic variance after a population bottleneck. *Trends Ecol. Evol.* **5** (7), 228–230. [https://doi.org/10.1016/0169-5347\(90\)90137-3](https://doi.org/10.1016/0169-5347(90)90137-3) (1990).
25. Gamblin, J., Marrec, L. & Olazcuaga, L. How bottlenecks shape adaptive potential: From theory and microbiology to conservation biology. (2024). <https://doi.org/10.32942/X2M608>
26. Nota, A., Bertolino, S., Tiralongo, F. & Santovito, A. Adaptation to bioinvasions: when does it occur? *Glob. Change Biol.* **30** (6), e17362. <https://doi.org/10.1111/gcb.17362> (2024).
27. Brugnone, F. et al. Atmospheric deposition around the industrial areas of Milazzo and Priolo Gargallo (sicily-Italy)—Part A: major ions. *Int. J. Environ. Res. Public Health.* **20** (5), 3898 (2023).
28. Calogero, G. S., Giuga, M., D'Urso, V., Ferrito, V. & Pappalardo, A. M. First report of mitochondrial DNA copy number variation in *Opius heydeni* (Insecta, Hymenoptera, Cicadellidae) from polluted and control sites. *Animals* **13** (11), 1793 (2023).
29. Martuzzi, M., Mitis, F., Biggeri, A., Terracini, B. & Bertollini, R. Ambiente e Stato Di salute Nella popolazione Delle Aree ad Alto Rischio Di Crisi ambientale in Italia. *Epidemiol. Prev.* **26**, 1–53 (2002).
30. Dlugosch, K. M. & Parker, I. M. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol. Ecol.* **17** (1), 431–449. <https://doi.org/10.1111/j.1365-294X.2007.03538.x> (2008).
31. Vitale, D. G. M. et al. Morphostructural analysis of the male reproductive system and DNA barcoding in *Balclutha brevis* Lindberg 1954 (Homoptera, Cicadellidae). *Micron* **79**, 36–45 (2015).
32. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30** (4), 772–780. <https://doi.org/10.1093/molbev/mst010> (2013).
33. Larsson, A. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30** (22), 3276–3278. <https://doi.org/10.1093/bioinformatics/btu531> (2014).
34. Allaire, J. RStudio: integrated development environments for R. *Boston MA.* **770** (394), 165–171 (2012).
35. Akts, C. & Akts, M. C. (2015). Package 'haplotypes'.
36. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods.* **9** (8), 772–772. <https://doi.org/10.1038/nmeth.2109> (2012).
37. Ronquist, F. et al. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61** (3), 539–542. <https://doi.org/10.1093/sysbio/sys029> (2012).
38. Kozlov, A. M., Darriba, D. & Stamatakis, A. (n.d.). RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference.
39. Rambaut, A. FigTree. Tree figure drawing tool. (2009). <http://tree.bio.ed.ac.uk/software/figtree/>
40. Bah, T. *Inkscape: Guide To a Vector Drawing Program* (prentice hall, 2011).
41. Kosakovsky Pond, S. L., Frost, S. D. W. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
42. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24** (8), 1586–1591. <https://doi.org/10.1093/molbev/msm088> (2007).
43. Yang, Z. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22** (4), 1107–1118. <https://doi.org/10.1093/molbev/msi097> (2005).
44. Murrell, B. et al. FUBAR: A Fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol. Evol.* **30** (5), 1196–1205. <https://doi.org/10.1093/molbev/mst030> (2013).
45. Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. W. GARD: A genetic algorithm for recombination detection. *Bioinformatics* **22** (24), 3096–3098. <https://doi.org/10.1093/bioinformatics/btl474> (2006).
46. Schwede, T. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **31** (13), 3381–3385. <https://doi.org/10.1093/nar/gkg520> (2003).
47. Jumper, J. et al. Highly accurate protein structure prediction with alphafold. *Nature* **596** (7873), 583–589 (2021).
48. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40** (D1), D370–D376. <https://doi.org/10.1093/nar/gkr703> (2012).
49. DeLano, W. L. The PyMOL molecular graphics system. (2002). <http://www.pymol.org/>.

Author contributions

Sampling: A.N. and F.T.; Experimental design: A.M.P. and V.F.; Laboratory processing: M.M., G.M., and G.S.C.; Sequencing: A.M.P.; Bioinformatics: M.M.; Figure preparation: M.M.; Manuscript writing and editing: M.M., F.T., A.M.P., A.N. and V.F.; All authors reviewed the manuscript equally.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-30855-z>.

Correspondence and requests for materials should be addressed to V.F. or A.M.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025