UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Biologia

CORSO DI DOTTORATO DI RICERCA IN BIOSCIENZE E BIOTECNOLOGIE

CURRICOLO: BIOLOGIA EVOLUZIONISTICA

CICLO XXIX

POPULATION GENOMICS APPROACH FOR THE STUDY OF CONNECTIVITY IN MARINE SPECIES

Coordinatore: Ch.mo Prof. Paolo Bernardi

Supervisore: Ch.mo Prof. Lorenzo Zane

Dottorando: Marta Paterno

To my Parents

Who taught me the importance

of challenging myself every day and stepping out of my comfort zone

TABLE OF CONTENTS

Thesis structure

The present thesis is a paper collection that describes and critically discusses the outcomes achieved throughout the 3-years PhD research activity. Under the supervision of Prof. Lorenzo Zane, I got involved in the European FP7 CoCoNET project, whose principal task was carrying out an exhaustive analysis of connectivity among existing and future marine protected areas (MPAs) in the Mediterranean and Black seas. Furthermore, I had the chance to spend a 6-months period at the University of Glasgow (Scotland) under the supervision of Dr. Martin Llewellyn, where I worked on a National Institutes for Health funded project regarding the Chagas disease in Latin America, whose aim was establishing the population genomic diversity of the vector (the local insect species *Rhodnius ecuadoriensis*) and the parasite (the single-celled kinetoplastid *Trypanosoma cruzi*). The thesis' structure traces the research activity I experienced in the 3-years PhD, so it's organized into two different-length sections:

- A main section related to the COCONET project [Part 1], entitled "MARINE CONNECTIVITY, SEASCAPE GENETICS and IMPLICATIONS for PROTECTED AREAS (MPAs)", of investigation and analysis of the populations connectivity pattern of 2 non-model species, the sea urchin *Paracentrotus lividus* and the Mediterranean mussel *Mytilus galloprovincialis* in the CoCoNET study areas;
- A minor section concerning the research activity carried out at the University of Glasgow [Part 2], entitled "LANDSCAPE GENETICS IN PARASITIC DISEASES", that deals with a pilot genomic study of the Chagas disease's insect vector *Rhodnius ecuadoriensis* in the area of disease transmission of Southern Ecuador.

The common point of these so different topics is the analytical framework: combining detailed environmental variation and genomics next-generation-sequencing data to obtain a better understanding of the biology and dispersal ability of the species under a seascape/landscape genetics approach.

Summary

MARINE CONNECTIVITY, SEASCAPE GENETICS and IMPLICATIONS for PROTECTED AREAS (MPAs)

Population connectivity is a keystone in a wide range of processes, including the shaping of population genetic structure, local and meta-population dynamics and the design of protected areas. Understanding the extent at which connectivity occurs is critical for the conservation of marine species and for the improvement of biological and ecological resources management. Genetic approaches are widely used as a powerful tool for the indirect inference of connectivity among populations of marine organisms, especially when the working-conditions makes the direct measure of dispersal pattern hard to quantify, such as the tracking of larval stage of benthic invertebrates. Despite their power, genetic isolation patterns can be difficult to explain, especially when the genetic differentiation is weak. The integration of genetic data with oceanographic feautures, the biology of species and the modelling of larval dispersal in a multidisciplinary approach, the so-called seascape genetics, can provide new insights into drivers shaping the spatial genetic variability of species. Worldwide, the anthropogenic activities have a strong negative impact to the marine ecosystem, causing coastal-habitat loss or fragmentation and the biodiversity depletion. Given the global extent of these phenomena, restoration of marine ecosystem is actually top conservation priority. Marine protected areas (MPAs), especially the no-take zones, have been increasingly identified as one of the most effective tool for conservation and management in reducing the alteration of marine ecosystem and mitigating the negative effect of human activities. The implementation of network of MPAs, where the individual protected areas can operate in synergy, could offer a greater protection at a wider spatial scale, leading more effectively the conservation goals than individual sites alone.

The European Commission funded CoCoNET project aimed to the identification of groups of putatively connected MPAs in the Mediterranean and the Black Seas, shifting from local to regional (Networks of MPAs) and basin (network of networks) scales. It adopted an interdisciplinary approach based on population genetics of 13 selected species, the modelling of the dispersal and the beta diversity.

In this thesis, 2 CoCoNET species were studied: the edible common sea urchin *Paracentrotus lividus* (Lamarck, 1816) and the Mediterranean mussel *Mytilus galloprovincialis* (Lamarck, 1819). The 2b-RAD protocol was successfully used to processing hundreds samples as well as for genotyping of hundreds SNPs, that were used to evaluating the genetic structure of these 2 species in the CoCoNET study areas.

• A seascape genomics approach was applied to explore connectivity patterns of the sea urchin *Paracentrotus lividus* in the Mediterranean pilot area: the potential larval connectivity and retention (Lagrangian simulations) and the realized connectivity (genomic analyses) were assessed and compared with a focus on the Adriatic and Ionian seas. A total of 275 samples from 10 population samples of Western and Central Mediterranean Sea were processed with the 2b-RAD protocol, and 1122 Single Nucleotide Polymorphisms (SNPs) were genotyped and used for population genomics analysis. Lagrangian simulations was performed with a biophysical model of larval dispersal focused on the Adriatic-Ionian basins and it was

integrated with features of the *P. lividus* larval phase (PLD, temperature constrain, reproductive timing). Genetic homogeneity was found among the eight population samples collected in the focal Adriatic-Ionian area, whereas a weak but significant differentiation was found with respect to two samples from the Western Mediterranean (France and Tunisia). This result was not affected by the few putative outlier loci identified in our dataset. Lagrangian simulations found a significant potential for larval exchange among the eight Adriatic-Ionian locations through a relatively persistent flux of propagules from each sites to its closest neighbor, mainly in a counterclockwise direction. Overall, these findings supported the hypothesis of connectivity of *P. lividus* populations in this area. We also compared our findings with previously published mitochondrial genetic data (cytochrome b gene) showing that our genome-scale SNPs dataset had a greater power in resolving the population structure of this species, providing solid and reliable indirect measures of connectivity, broadly in line with the predictions of Lagrangian simulations.

• The population genomics structure of the mussel *Mytilus galloprovincialis* was explored at different geographic scales (local and wide) in the Mediterranean and Black seas. The application of a genome-wide approach based on single nucleotide polymorphism (SNP) markers, a wide sampling design and a good sample size, allowed to overcome the previous limitations providing the first reliable biogeographic data concerning this species throughout its native range. A total of 19 population samples were analysed, 10 from the Mediterranean Sea and 9 from the Black Sea. To obtain the best resolution for the wide and the local scale, we obtained and investigated 3 different datasets: 1) the overall dataset (461 individuals, 512 polymorphic loci); 2) the Adriatic-Ionian dataset (201 individuals, 811 polymorphic loci) and 3) the Black Sea's dataset (228 individuals, 998 polymorphic loci). A marked structuring between the Mediterranean and Blacks Seas was found, revealing that the Mediterranean and Black seas are genetically distinct, likely due to oceanographic constrains. In the Black Sea, a completely lack of genetic structure was detected, where the currents pattern and the long pelagic larval duration of this species probably ensure a good connection among the sampled locations. On the other hand, in the Mediterranean Sea, no signal of differentiation was found between the Western and Central, whereas the Central Mediterranean (the Adriatic-Ionian population samples) showed a weak genetic heterogeneity, that underlined an East-West differentiation in the Central Adriatic Sea, and a general homogeneity in the Southern Adriatic and Ionian seas among the Italian, Greek and Albanian localities. This pattern of genetic differentiation likely depends on the Adriatic gyral circulation, that prevents or allows the exchange of larvae, acting as barrier among some localities and determing a certain level of self retention in others.

Overall, connectivity data obtained so far in the CoCoNET framework revealed a heterogeneous mosaic of species-specific connectivity patterns throughout the targeted areas in Mediterranean and Black seas, ranging from the total absence of genetic differentiation (*P. lividus*) to the strong genetic structure even at a local scale (in the habitat former *P. oceanica*). Overall, the heterogeneity of connectivity patterns depends on species-specific peculiarities in relation to environmental features, and connectivity results so far

obtained warns against any generalization in defining conservation units based on the results of the single species. The wide heterogeneity of species-specific connectivity outcomes suggests that the implementation of ecological effective networks of MPAs should standing on meta-community information across a range of spatial scales and a representative set of species to obtain a "collection of networks of genetic variation of all species within a community".

LANDSCAPE GENETICS IN PARASITIC DISEASES

Neglected Tropical Diseases (NTDs) are a group of parasitic infectious diseases affecting worldwide approximately 2.7 billion people living in poverty, especially in the tropical areas. The NTDs pathogens have often a complex life-cycle, involving at least one host, including humans. A great heterogeneity exists in the dispersal ability, distribution and transmission routes of parasitic species, that are strongly influenced by ecological settings and landscape variation. Understanding how abiotic and biotic variables interact and affect the movements and distribution of parasites, hosts and vector species is key for predicting and controlling the disease spread. *Trypanosoma cruzi* is the most important parasite in Latin America, the infectious agent of Chagas disease, a NTD that caused around 14000 deaths annually. In Ecuador, where the Chagas disease is endemic, one of main insect vectors is the kissin bug *Rhodnius ecuadoriensis*.

In this thesis we worked on a pilot study regarding the vector *Rhodnius ecuadoriensis* based on 20 specimens collected from four communities in Ecuador, with the aims of i) testings the 2b-RAD protocol by assessing and comparing the effectiveness of three type IIB Restriction Enzymes: ii) generating several hundred SNPs markers for this species and iii) performing preliminary genomic analysis to test the efficacy of the identified SNP markers. The three type IIB Restriction Enzymes (AlfI, BcgI, and CspCI) resulted to be effective for processing *Rhodnius ecuadoriensis* genomic DNA samples and, as expected, the enzyme showing the most abundance of cutting sites (estimated via an *in silico* digestion of the *Rhodnius prolixus* genome) provided the smaller number fo usable molecular markers (AlfI). CspCI enzyme produced the largest amount of polymorphic markers whatever the read depth, thus evidencing its good experimental performance in *R. ecuadoriensis* and likely to other *Rhodnius* sp. Vectors. Preliminary population genomic analyses were performed with BgcI and CspCI datasets, and a signal of structuring among populations was detected. Specifically, Bayesian clustering detected two distinct clusters without previous location information, one of them was the most distant population (Bejuco) from the others 3 from Loja region, that clustered together. Our findings suggest that 2b-RAD genotyping is both a cost-effective and methodologically simple approach for generating high resolution genomic data for Chagas disease vectors with the power to distinguish between different vector populations at epidemiologically relevant scales.

Riassunto

CONNETTIVITA' MARINA, SEASCAPE GENETICS E IMPLICAZIONI PER LE AREE MARINE PROTETTE

La connettività di popolazione è la chiave di volta di un'ampia varietà di processi, tra i quali la determinazione della struttura genetica di popolazione delle specie, le dinamiche di popolazioni locali e meta-popolazioni e la progettazione di aree protette. Capire e definire le modalità con cui le popolazioni sono connesse e a che scala geografica la connessione avviene, è critico per la conservazione delle specie marine e per migliorare la gestione delle risorse marine economiche e biologiche. Gli approcci genetici sono ampiamente utilizzati come mezzo efficace per la stima indiretta della connettività tra popolazioni di organismi marini, specialmente quando le condizioni di lavoro rendono difficile lo studio diretto dei *pattern* di dispersione, come nel caso del *tracking* delle larve negli invertebrati bentonici. Nonostante il potere della genetica, i *pattern* di isolamento rilevati possono risultare di difficile interpretazione, soprattutto quando il differenziamento genetico osservato è debole. Integrare i dati genetici con le caratteristiche oceanografiche dell'area di studio, dati riguardanti la biologia delle specie e il *modelling* della dispersione larvale in un approccio multidisciplinare chiamato *seascape genetics*, può chiarire quali siano i fattori che determinano la distribuzione spaziale della variabilità genetica delle specie.

In tutto il mondo, le attività antropiche hanno effetti estremamente negativi sull'ecosistema marino, causando la perdita o la frammentazione degli habitat costieri e il depauperamento della biodiversità ad essi associata. Data l'estensione globale di questi fenomeni, il ripristino dell'ecosistema marino è una delle principali priorità target nell'ambito della conservazione. Le aree marine protette (MPAs) sono state riconosciute come uno dei mezzi più efficaci per conservare e gestire l'intero ecosistema marino, in quanto sono in grado di ridurre e mitigare gli effetti negativi causati delle attività umane. L'implementazione di *network* di MPAs, all'interno dei quali le singole aree marine protette possono agire in sinergia, potrebbe offrire una maggior protezione su ampia scala spaziale, agendo in modo più efficace rispetto a quanto possano fare le singole aree da sole.

Il progetto CoCoNET, finanziato dalla Commissione Europea, è un progetto che mirava all'identificazione di gruppi di MPAs putativamente connesse nel Mediterraneo e nel Mar Nero, sia su scala locale sia considerando l'intero bacino. Il progetto CoCoNET si proponeva di adottare un approccio multidisciplinare basato sulla genetica di popolazione di 13 specie, il *modelling* della dispersione larvale e la *beta diversity*.

In questa tesi sono state studiate 2 specie del progetto CoCoNET, il riccio di mare *Paracentrotus lividus* (Lamarck, 1816) e il mitilo del Mediterraneo *Mytilus galloprovincialis* (Lamarck, 1819). Per il processamento di centinaia di campioni e il *genotyping* di altrettanti polimorfismi a singolo nucleotide (SNPs) utilizzati per studiare la struttura genetica di popolazione delle 2 specie, è stato utilizzato il protocollo 2b-RAD.

• Per il riccio di mare *P. lividus* è stato utilizzato un approccio di *seascape genomics* nel Mediterraneo centrale: sono stati stimati e confrontati il potenziale di dispersione larvale e di *retention* (ottenuti con simulazioni di Lagrange) e la connettività realizzata (ottenuta con i dati genetici) con focus a livello

dell'Adriatico e dello Ionio. Sono stati processati un totale di 275 individui da 10 campioni di popolazione provenienti dal Mediterraneo occidentale e centrale, e 1122 SNPs sonno stati genotipizzati ed utilizzati per le analisi di genomica di popolazione. Le simulazioni di Lagrange sono state condotte utilizzando un modello biofisico di dispersione larvale focalizzato nel bacino Adriatico-Ionico, che è stato integrato con caratteristiche dello stadio larvale di *P. lividus* (durata delle larve, influenza della temnperatura, periodo riproduttivo). Gli 8 campioni di popolazione campionati in Adriatico-Ionio sono risultati essere geneticamente omogenei, mentre è stato rilevato un lieve ma significativo differenziamento di queste 8 rispetto ai 2 campioni di popolazione provenienti dal Mediterraneo occidentale (Francia e Tunisia). Questo risultato non è stato influenzato dalla presenza di alcuni loci probabilmente sotto selezione direzionale. Le simulazioni di Lagrange hanno predetto uno scambio larvale potenziale tra gli 8 campioni di popolazione Adritico-Ionici attraverso un flusso di propaguli relativamente persistente da ogni sito, prevalentemente in direzione oraria. Complessivamente, i risultati ottenuti dalla genetica di popolazione e dalle simulazioni supportano ampiamente l'ipotesi di una buona connessione tra le popolazioni di *P. lividus* in quest'area.

• La struttura genomica di popolazione del mitilo *M. galloprovincialis* è stata esplorata su diverse scale geografiche nel Mediterraneo e Mar Nero. L'utilizzo di un approccio *genome-wide* basato su SNPs, un ampio campionamento e una buona dimensione dei campioni di popolazione hanno permesso di superare le limitazioni finora incontrate nell'ottenere dati biogeografici affidabili per questa specie all'interno del suo range di distribuziona nativo. Sono state analizzati 19 campioni di popolazione, 10 provenienti dal Mar Mediterraneo e 9 dal Mar Nero. Al fine di ottenere il miglior potere di risoluzione possibile sia su ampia scala sia su scala ridotta, sono stati utilizzati per le analisi 3 diversi dataset: 1) un dataset complessivo (461 individui, 512 loci); 2) un dataset adriatico-ionico (201 individui, 811 loci) e 3) un dataset per il Mar Nero (228 individui, 998 loci). È stato evidenziato un marcato differenziamento genetico tra il Mar Mediterraneo e il Mar Nero, probabilmente a causa di *costrain* oceoanografici. Nel Mar Nero è evidente una totale mancanza di struttura, probabilmente dovuta al pattern delle correnti e alla durata dello stadio larvale che permettono una buona connessione tra le popolazioni campionate. Dall'altro lato invece, non è stato notato differenziamento tra il Mediterraneo occidentale e quello centrale, mentre un certo grado di eterogeneità genetica è stato rilevata tra le popolazioni adriatico-ioniche, con un differenziamento est-ovest nell'Adriatico centrale e una relativa omogeneità tra le popolazone dell'Adriatico meridionale e dello Ionio. Il *pattern* di differenziamento osservato dipende probabilmente dalla presenza di *gyres* minori in aggiunta alla circolazione adriatica principale, che può favorire o impedire lo scambio di propaguli, fungendo da barriera al flusso genico tra alcune località e favorendo la *self-retention* in altre.

Complessivamente, i risultati ottenuti finora nel progetto CoCoNET hanno messo in luce un mosaico eterogeneo di *pattern* di connettività specifici per le diverse specie, che vanno dalla totale assenza di differenziamento come nel caso del riccio, alla forte struttura come nel caso dell'*habitat former Posidonia oceanica.* Questa eterogeneità può dipendere da peculiarità specie-specifiche messe in relazione a

caratteristiche ambientali, e i dati di connettività finora ottenuti sconsigliano qualsiasi forma di generalizzazione basandosi sui risultati ottentui in una singola specie. Il quadro che sta emergendo suggerisce che la progettazione di network di MPAs ecologicamente efficaci debbano basarsi su informazioni ottenute a livello di comunità considerando un ampio *range* spaziale, e un set rappresentativo delle specie che vivono nell'area *target*.

LANDSCAPE GENETICS NELLE PARASSITOSI

Le malattie tropicali sono un gruppo di malattie infettive causate da parassiti che infettano quasi 3 milioni di persone in tutto il mondo, specialmente nei paesi poveri delle zone tropicali. I patogeni che causano questa malattie hanno solitamente un ciclo vitale complesso, che coinvolge almeno un ospite, compresi gli esseri umani. Esiste una grande variabilità nelle capacità di dispersione, nella distribuzione e nelle vie di trasmisisone delle specie parassite. Questa variabilità è fortemente influenzata dalle caratteristiche ambientali e del paesaggio. Capire come le variabili biotiche e abiotiche interagiscono e influenzano la distribuzione del parassita, dei loro ospiti e delle specie vettore è un fattore chiave per la predizione ed il controllo della diffusione di queste malattie. Il *Tripanozoma cruzi* è il parassita più importante del Sud America ed è l'agente infettivo del morbo di Chagas, una malattia tropicale endemica in molte regione dell'Ecuador dove uno dei principali vettori è la cimice ematofaga *Rhodnius ecuadoriensis*.

In questa tesi è stato prodotto uno studio pilota sul vettore *R. ecuadoriensis* utilizzando 20 individui campionati in 4 comunità dell'Ecuador, con lo scopo di i) testare il protocollo 2b-RAD utilizzando e confrontanto l'efficacia di 3 diversi enzimi di restrizione di tipo IIb, ii) ottenere un numero di marcatori SNPs sufficiente per iii) fornire delle analisi preliminari di genetica di popolazione per questa specie. I 3 enzimi testati (AlfI, BcgI, and CspCI) sono risultati efficaci nel processare il DNA genomico di questa specie e, come era atteso, l'enzima con il maggior numero di siti di riconosciemento (stimato tramite una restrizione *in silico* del genoma di *Rhodnius prolixus*) ha prodotto il minor numero di loci utilizzabili (AlfI). L'enzima CspCI ha prodotto il maggior numero di loci polimorfici a qualsiasi *read depth* considerata evidenziando così l'ottima performance di questo protocollo in questa specie e presumibilmente anche in altre specie appartenenti allo stesso genere. Le analisi preliminari di genetica di popolazione sono state condotte utilizzando i dataset ottenuti dagli enzimi BcgI e CspCI che sono stati in grade di evidenziare un certo grado di struttura genetica tra I 4 campioni di popolazione. Un *clustering* bayesiano, condotto senza l'informazione a priori riguardante le località, ha identificato la presenza di 2 *cluster* geneticamente distinti: uno rappresentato dalla popolazione geograficamente più distante rispetto alle altre (Bejuco), l'altro invece formato dalle altre 3 popolazioni che provenivano tutte dalla regione di Loja. Questi risultati evidenziano come il protocollo 2b-RAD sia una metodologia semplice ed efficace nel generare marcatori su scala genomica in uno dei principali vettori del morbo di Chagas, e come permetta di ottenere informazioni sulla struttura di popolazione di questa specie ad una scala rilevante a livello epidemiologico.

Part 1

MARINE CONNECTIVITY, SEASCAPE GENETICS and IMPLICATIONS for PROTECTED AREAS (MPAs)

CHAPTER 1

Background and aims

Background and aims

Population connectivity

Talking about connectivity is the trend of the moment in both ecology and evolution of species (Crooks and Sanjayan 2006; Weersting and Toonen 2009; Carson et al. 2011; Lòpez-Duarte et al. 2012; Kool et al. 2013; Melià et al. 2016; Selkoe et al. 2016). Connectivity is essential in a wide range of processes, including the shaping of population genetic structure, local and meta-population dynamics, the controlling of invasive species, the resiliency to human exploitation and climate changes, and the design of protected areas (Hastings and Harrison, 1994; Cowen et al. 2007; Hulme 2009; Munday et al. 2009; Weersting and Toonen 2009; Kool et al. 2013; Grorud-Colvert et al. 2014; Puckett et al. 2014). Understanding the extent at which population connectivity occurs is key for the conservation of marine species and for the improvement of biological and ecological resources management (Moritz 1994; Palumbi et al. 2003; Cowen et al. 2006; Weersing and Toonen 2009; Carson et al. 2011).

Over the past decades, a variety of different approaches have been employed to yielding new insights into connectivity trends of species. Among the techniques directly measuring the connectivity, it's possible to mention the classic mark-recapture method, widely used for terrestrial organisms (Seber 1982; Webster et al. 2002), the analysis of the chemical signatures of otoliths and statoliths in fish invertebrates respectively (Campana 1999; Thorrold et al. 2007), the artificial tagging in marine species (Almany et al. 2007) and the kinship analysis for the direct estimates of dispersal distances (reviewed in Selkoe et al. 2016). Moreover, indirect estimates from molecular markers analysis (Hellberg et al. 2002; Palumbi 2003), oceanography patterns (Lee et al. 1994) and the biology of larvae (Emlet et al. 1987) have been extensively used.

Not less than 70% of marine invertebrates has a pelagic phase, with a planktonic larval stage that can be transported by currents from spawning sites to the open sea over long distances towards suitable habitats for the settlement and development into a sedentary or sessile adult stage (Mileikovsky 1971; Levin 2006). Given that connectivity is defined as "the extent to which populations in different parts of a species' range are linked by the exchange of larvae, recruits, juveniles, or adults" (Palumbi 2003), populations connectivity in these species is primarily driven by the dispersal ability of the larval stage (Gilg and Hilbish 2003; D'Aloia et al. 2015). Ideally, the dispersal can be quantified by tracking propagules and their movements, but the rapid dilution of larvae over time and space (Cowen and Sponaugle, 2009), their small size, the high fecundity of marine species and the high mortality rate of the early life stages (Thorrold et al. 2002), make the assessment of connectivity in marine environment much more complex.

Population genetics is a valuable tool for the indirect study of connectivity, thus the distribution of the genetic variance within species reflects the rate of exchange between populations (Palumbi 2003) as well as the drift within populations. The pelagic larval duration (PLD) has been traditionally considered a proxy for the realized dispersal, following the conventional, long-standing and intuitive hypothesis that long PLD

confers a great dispersal capability, and thus a robust rate of gene flow between populations, that has been fully supported for a long time (Jablonski 1986; Doherty et al. 1995; Bohonak 1999; Shanks et al. 2003; Kinlan et al. 2005; Lester et al. 2007). It was commonly assumed that the dispersal capacity of species declines with the distance from the source, reducing the number of migrants and increasing the genetic differentiation between populations (Wright 1931). This hypothesis has been recently criticized and rejected by a number of studies (revised in Weersing and Toonen 2009) that found a weak correlation between the average duration of larval stage and the genetic differentiation (measured by F_{ST} indexes) in different marine taxa. In several cases, strong genetic partitioning was found despite a long PLD (Todd 1998; Bowen et al. 2006), suggesting that marine populations can be less connected than expected. Limited exchange between marine populations, even at small spatial scales, may be explained by a number of biophysical mechanisms related to the biology of species, ecological and environmental factors, and the genetic analysis conducted (Taylor and Hellberg 2003). Specifically, the duration of the larval stage is a plastic trait that can varies in response to external parameters like salinity, temperature or food availability (Pechenik 1990; Woodson and McManus 2007) and the larval behavior, in terms of swimming ability, actively effects the position of larvae within the water masses (Sponaugle et al. 2002). In the marine environment, Euclidean distance between locations is often a poor predictor of realized dispersal distance, because, even when there are no evident barriers to the gene flow, the complexity of water circulation patterns (eddies and gyres) can determine the retention and/or diffusion of pelagic larvae at very different spatial scales (Ruzzante et al. 1998; Oalumbi 2003; Cowen and Sponaugle 2009, White et al. 2010). Moreover, the experimental approach, including the sampling design and expecially the choice of the genetic marker class for the inference of the differentiation, has a significant effect in detecting genetic variability. Thus, values of genetic structuring estimated from different markers, with different features (polymorphisms level, inheritance and so on), are not often equivalent and comparable (Weersing and Toonen 2009). For example, mitochondrial DNA markers generally found a much higher genetic differentiation than nuclear loci, probably due to the reduction of the population size because of the uniparental inheritance (if compared to the biparental one) that leads these markers to experience an increased rate of genetic drift (Wright 1931; Ballard and Whitlock 2004; Weersing and Toonen 2009; Paterno et al. 2017). In these working-conditions, enlightening magnitude and directions of larval dispersal and quantifying the exchange between populations of marine species is still a major present-day challenge.

From population genetics to genomics

Genetic approaches are commonly used as a powerful source of information for the inference of connectivity among populations of marine organisms, especially when the direct measure of dispersal pattern would be otherwise hard to quantify, such as the tracking of larval stage of benthic invertebrates

(Cowen and Sponaugle 2009; Hellberg et al. 2002; Palumbi, 2003; Thorrold et al., 2002; Broquet and Petit 2009; Lowe and Allendorf 2010; Selkoe and Toonen 2011).

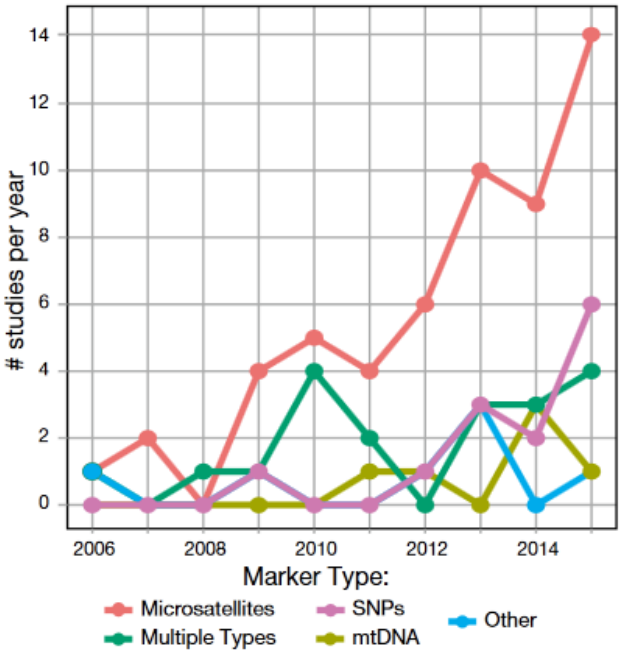


Figure 1. Trends in marker usa over time (Selkoe et al. 2016). Microsatellites are the most common marker type used, SNPs studies are rising. Multiple Types: studies that use more than one marker type, with the most common combination being microsatellites mtDNA. Other: marker types used infrequently (e.g. inter simple sequence repeats, allozymes, single nuclear sequences).

Traditionally, studies in the fields of conservation and population genetics have relied on a few molecular markers (Fig. 1), including allozymes (Bader 1998), RFLPs (Botstein et al. 1980), AFLPs (Vos et al. 1995), and microsatellites (Jarne and Lagoda 1996). Microsatellites in particular, became very popular due to their hypervariability and are currently used in the majority of the recent studies. Despite their success, microsatellites development is costly and time-consuming, and requires *a priori* genetic information (primer design). In addition, microsatellites are difficult to model, because we are not completely aware of their mutational model, thus limiting the power of many recent analytical progresses. Finally, microsatellites cover a very limited portion of the genome and are considered putatively neutral, limiting their ability to investigate adaptive genetic variation or evolutionary events like resistance to pathogens (Bonneaud et al. 2011) and local adaptation to a new habitat (Narum et al. 2010).

"The ideal molecular approach for population genomics should uncover hundreds of polymorphic markers that cover the entire genome in a single, simple and reliable experiment. Unfortunately, at present there is no such approach" (Luikart et al. 2003). Now, we are living the advent of next-generation sequencing (NGS) which hold the promise to reach this ideal approach. Even if the complete genome sequencing at the population level is still not affordable, NGS-based approaches allow to sampling cheaply, randomly and densely the genome for discovering, sequencing and genotyping hundreds up thousands of markers, even when no a priori genetic information is available. Currently, population studies using NGS techniques often

stand on the use of restriction endonucleases and involve key steps (genomic DNA digestion, selection and sequencing of the resulting restriction fragments) to obtaining a reduced representation of the target genome (Davey et al. 2011; Narum et al. 2013). The restriction-site-associated DNA sequencing (RAD-seq; Baird et al. 2008; Davey and Blaxter 2010) is a class of the NGS techniques widely used as a powerful tool in a variety of fields, from genetic differentiation to the construction of linkage maps and adaptation studies (Baird et al. 2008; Hohenlohe et al. 2010; Baxter et al. 2011; Etter et al. 2011; Andrews et al. 2016). All RAD-seq protocols start by digesting genomic DNA sample with one or more chosen restriction enzymes, followed by the ligation of individual-specific barcodes to the restriction fragments, that will be typically sequenced on an Illumina Genome Analyzer platform. The single nucleotide polymorphisms (SNPs) identified and genotyped in the sequenced restriction fragments can be used as genomic markers (Xu et al. 2016). The possibility to address questions on a wide range of evolutionary processes acting at the genome-scale (Li et al. 2012), and the power, simplicity and the relative low-cost of the NGS techniques, have promoted the recent and rapid growth of the SNP-based genetic studies (Fig. 1; Selkoe et al. 2016).

The RAD-seq protocols differ in the number and class of restriction enzymes employed: original RAD (one enzyme; Baird et al. 2008), 2b-RAD (one IIB enzyme; Wang et al. 2012), genotype-by-sequencung or GBS (one or 2 enzymes; Elshire et al. 2011; Sonah et al. 2013), ezRAD (one or more enzyme; Toonen et al. 2013) and ddRAD (2 enzymes; Peterson et al. 2012). Here, among all, the 2b-RAD protocol was preferred (Fig. 2).

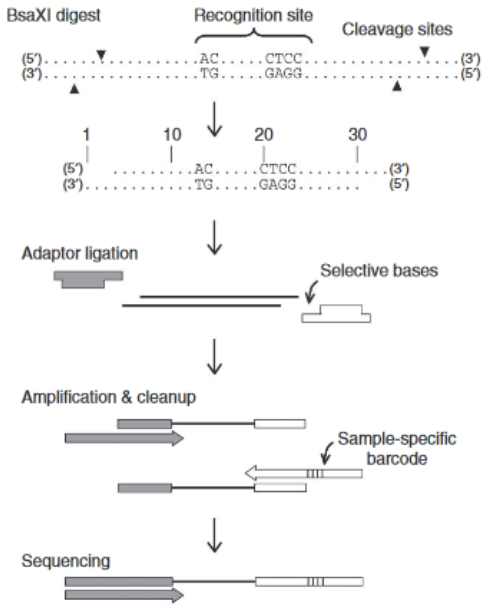


Figure 2. Preparation and sequencing of 2b-RAD tags (Wang et al. 2014). Restriction digestion (BsaXI) of genomic DNA, cohesive-end ligation of partially double-stranded adaptors with compatible (NNN) overhangs, and incorporation of barcodes for multiplex sequencing by PCR.

This protocol relies on the use of IIB restriction enzymes that have the peculiarity to cleave genomic DNA upstream and downstream the recognition site (Marshall et al. 2010), producing restriction tags of uniform length (33-36 bp). This feature, by contrast with other RAD protocols, removes the restriction fragments'

size selection and, at the same time, the preferential amplification of short fragments over longer fragments. Moreover, 2b-RAD produces sequence data at all the enzyme cutting sites, providing a higher number of suitable loci and consistency among different libraries. This protocol allows to process a large number of (barcoded) DNA samples in parallel, and it results to be accurate in SNPs identification and genotyping both with a reference genome or in *de novo* analysis (Pauletto et al. 2016; Pecoraro et al. 2016). In the present thesis, the 2b-RAD protocol was successfully used for processing a great number of genomic DNA samples from 3 different species: the sea urchin *Paracentrotus lividus*, the Mediterranean mussel *Mytilus galloprovincialis* and the insect vector *Rhodnius ecuadoriensis*. This approach leaded to the identification of hundreds of SNP markers with suitable polymorphism level and read depth for all 3 species under examination, allowing a good power of resolution in defining their population genetic structure at a very different spatial scale and without a reference genome, holding the authors' great promise of applicability "for studies of ecology and evolution in diverse species" (Wang et al. 2012).

Seascape genetics

As above-mentioned, characterizing the genetic structure of marine species has many challenges, including the collection of samples, the poor knowledge of the interactions between biotic and abiotic factors affecting the realized dispersal of species, and the development of suitable molecular markers. Despite that genetic methods are one of the most powerful tool in the hands of scientists for inferring connectivity in marine environment, genetic isolation patterns, especially when the genetic signal is weak, are often hard to explain (Waples 1998). Integrating oceanography, biology of species (larval behavior) and modelling of dispersal with genetics data in a multidisciplinary approach, can provide new insights into drivers shaping the spatial genetic variability of species. This integrated approach is what seascape genetics stands on (Galindo et al. 2006; Hansen-Hemmer-Hansen 2007; Selkoe et al 2008; White et al. 2010). Seascape genetics is a rapidly growning tenth-years-old discipline, with many studies highlighting the strength of integrating genetic data and oceanographic features in detailing the population genetic structure of marine species from different taxonomic groups, from invertebrates (Gilg and Hilbish 2003; Hohenlohe 2004; Kenchington et al. 2006; Dupont et al. 2007; White et al. 2010; Schiavina et al. 2014; Paterno et al. 2017; Marino et al. in review) to fish (Di Franco et al. 2012; Pujolar et al. 2013). For example, Gilg and Hilbish (2003) combined genetics, realized larval dispersal and a coastal circulation model the southwest England to determine the extent of dispersal of larvae of *Mytilus edulis* and *M. galloprovincialis* from a hybrid zone to an eastern putative pure *M. edulis* genetic region and to a putative pure *M. galloprovincialis* region in north Cornwall. The authors found a match between genetic data and oceanographic results, with the hydrodinamic model that accurately predicted the general pattern and the scale of larval dispersal between the selected regions (Gilg and Hilbish 2003). Schiavina et al. (2014) investigated the population structure of

the Mediterranean shore crab *Carcinus aestuarii* in the central Mediterranean (Adriatic Sea) coupling the assessment of the genetic differentiation (microsatellites markers) and estimates of potential larval connectivity obtained by a physical–biological model describing the water circulation pattern of the area with the inclusion of those biological traits that can affect *C. aestuarii*'s dispersal ability. The authors compared the results from the 2 methods gaining support to the view that the genetic differentiation of *C. aestuarii* in the central Mediterranean reflects the oceanographic subdivision of the Adriatic Sea into 3 sub-basins, with circulation patterns allowing the exchange of larvae through permanent and ephemeral connections (Schiavina et al. 2014). Generally, comparing genetic differentiation obtained from molecular markers and results from physical-biological model predicting the potential for larval dispersal within the study area, underlined the importance of currents for shaping genetic structuring in marine species by presenting or allowing the connection between populations.

Marine protected areas (MPAs)

Worldwide, the anthropogenic activities have a strong negative impact to the marine ecosystem, such the coastal-habitat loss or fragmentation and the biodiversity depletion due to fishing and polluting pressures, the urbanization and the invasion of alien species (Jackson et al. 2001; Boudouresque et al. 2005; Lotze et al. 2006; Airoldi and Beck 2007; Claudet and Fraschetti 2010) and climate-induced stressors (Lejeusne et al. 2010). In addition to the ecological damage, we experience an economic loss due to the decrease of goods and services in response to the overexploitation of marine resources (Worm et al. 2006; Halpern et a. 2007; Fenberg et al. 2012). In this global scenario, the restoration and conservation of marine biodiversity and environment are top priorities.

Marine protected areas (MPAs), especially the no-take zones, have been increasingly recognized as one of the most effective tool for conservation and management in reducing the alteration of marine ecosystems and mitigating the negative anthropic impacts (Gaines et al. 2010; Gabrié C. et al. 2012). The effectiveness of MPAs in fulfilling ecological restoration goals has been studied, highlighting the ability to i) provide refuges for exploited and endangered species, ii) allow the re-establishment of trophic relationships and community interactions and iii) act as a source of larvae, juveniles and adults to surrounding unprotected waters (Sale et al. 2005; Guidetti and Sala 2007; Claudet et al. 2008; Fraschetti et al. 2011; Di Franco et al. 2012; Fenberg et al. 2012). This fast-degenerating scenario calls for protection and conservation at a wide geographical scale, thus a local management based on isolated marine protected areas could not be sufficient. The implementation of network of MPAs, where the individual protected areas can cooperatively operate in synergy, could offer a greater protection at a wider spatial scale, leading more effectively the conservation goals than individual sites alone (Wood et al. 2008; Wabnitz et al. 2010; Fenberg et al. 2012; Grorud-Colvert et al. 2014). The World Summit on Sustainable Development (WSSW) together with the Fifth World Parks Congress (WPC) stressed the importance and especially the need to set up effective and

representative networks of MPAs by 2012. But, because of the poor knowledge about the extent of the most of habitats, of the potential connectivity capability of species, the realized dispersal in and out the reserve and the rates of exchange between populations, we are still so far from this goal.

The assessment of whether protected sites are ecologically connected or not, by means of the study of genetic variability distribution of species, can provide a useful tool to determining if existing or future protected area are operating or could operate as an effective network or not (Fenberg et al. 2012). The study of connectivity under a seascape genetics/genomics framework can be even more powerful, helping to explain the genetic distribution of the species in the light of the interaction between oceanography and species-specific dispersal ability (Di Franco et al. 2012; Pujolar et al. 2013; Paterno et al. 2017).

The CoCoNET project

CoCoNet (Towards Coast to Coast NETworks of Marine Protected Areas Coupled with Sea-based Wind Energy Potential) was a 4-years project made with funding from the European Commission's Seventh Framework Programme (FP7/2007-2013; grant agreement n° 287844) and involved 37 partners from 22 countries of three continents (for more details see the CoCoNET website *http://www.coconet-fp7.eu/*).

The CoCoNet project's goals were i) providing solid guidelines for the establishment MPAs networks in the Mediterranean and Black Seas based on connectivity; and ii) producing a 'wind atlas' to assess the feasibility of Offshore Wind Farms (OWFs) in the same study areas. The main goal of CoCoNet was the development a deep knowledge for the establishment of MPAs networks and assessing the feasibility of OWFs in the Mediterranean and Black Seas.

The CoCoNET aim involving MPAs networks is linked to my PhD's activity. CoCoNET aimed to the identification of groups of putatively connected MPAs in the Mediterranean and the Black seas (Fig. 3), shifting from local (single MPA) to regional (Networks of MPAs) and basin (network of networks) scales.

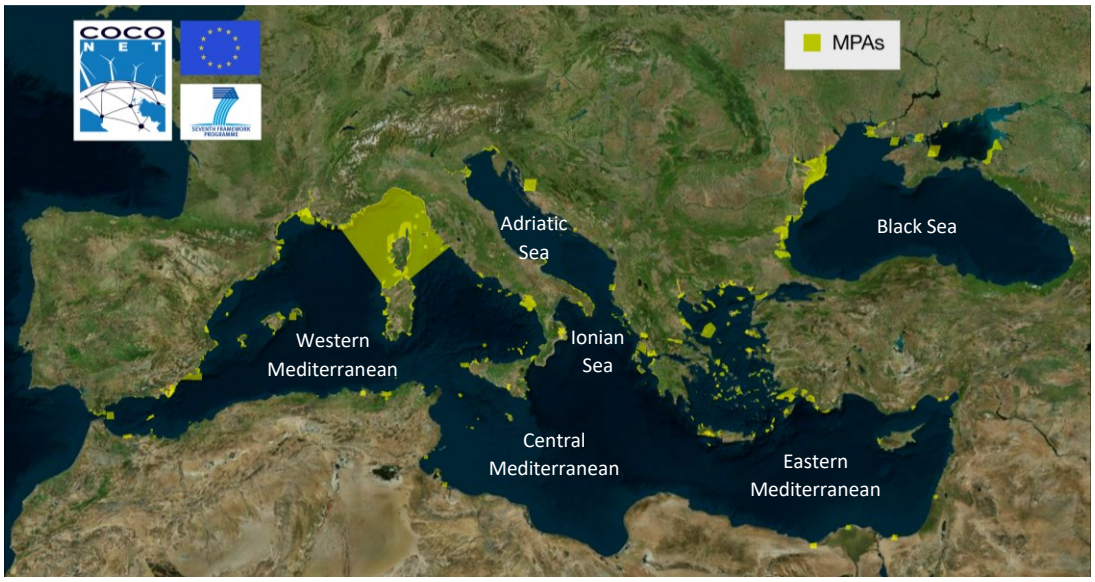


Figure 3. Representation of Marine Protected Areas (MPAs) in the Mediterranean and Black seas. Source: *http://www.coconet-fp7.eu/*

The novelty and the power of this project lied on the application of an interdisciplinary approach based on population genetics, the modelling of marine species dispersal and the beta diversity. The genetic connectivity was assessed on the population structure of 13 selected species of different taxonomic groups with different ecological roles and dispersal mechanisms: *Cystoseira barbata* (brown alga), *Zostera noltei* (seagrass), *Mytilus galloprovincialis* (bivalve mollusc), *Gibbula divaricata* (gastropod mollusc), *Cyclope neritea* (gastropod mollusc), *Pachygrapsus marmoratus* (crustancean), *Scorpaena porcus* (bony fish), *Symphodus tinca* (bony fish), *Posidonia oceanica* (seagrass), *Desmophyllum dianthus* (scleractinian coral), *Paracentrotus lividus* (sea urchin), *Cladocora caespitosa* (scleractinian coral), *Donacilla cornea* (bivalve mollusc). These species were sampled within 2 pilot areas in the Adriatic-Ionian basins (8 sites) and in the Black Sea (8 sites) at localities chosen accordingly to the knowledge of current oceanographic patterns, the habitat distributions of the target species and the location of effective and future MPAs. The identification of physical and biological connections will clear those processes driving the patterns of biodiversity distribution allowing the enhancement of the policies for an effective environmental management, also checking if the existing MPAs are sufficient for an ecological networking and suggesting how to design further protection schemes based on effective exchanges between protected areas.

Thesis' goals

The goal of my PhD research project was the evaluation of the connectivity patterns in the Mediterranean and the Black seas in the CoCoNET framework. To this aim, the edible common sea urchin *Paracentrotus lividus* (Lamarck, 1816) and the Mediterranean mussel *Mytilus galloprovincialis* (Lamarck, 1819) were selected as target species for the genetic assessment of the connection between the chosen localities.

Probably these 2 species alone are not the best candidates to outline the connectivity among protected areas, because the long-dispersing trait could represent a limit in the detection of connectivity patterns. However, the facility of samples collection, the wide distribution thoughout the study areas, the important ecological role and the possibility to compare the connectivity outcomes from different species under the multi-species approach of the CoCoNET project, made the sea urchin *P. lividus* and the mussel *M. galloprovincialis* good targets for the connectivity investigation of this thesis.

For both species, the RADseq protocol 2b-RAD was successfully used processing hundreds samples as well as for genotyping of hundreds SNPs, that were used to evaluating their genetic structure in the CoCoNET study areas. For the sea urchin, it was also possible estimating the dispersal potential of the larval phase via Lagrangian simulations across the sampling sites selected for the genetics. In detail:

- A seascape genomics approach was applied to explore connectivity patterns of the sea urchin *Paracentrotus lividus* in the Mediterranean pilot area: the potential larval connectivity and retention (Lagrangian simulations) and the realized connectivity (genomic analyses) were assessed and compared

with a focus on the Adriatic and Ionian seas. We also compared our findings with previously published genetic data based on mitochondrial markers (cytochrome b gene) showing how our genome-scale SNPs dataset had a greater power in resolving the population structure of this species, providing solid and reliable indirect measures of connectivity, broadly in line with the predictions of Lagrangian simulations.

- The population genomics structure of the mussel *Mytilus galloprovincialis* was explored at different geographic scale (local and wide) in the Mediterranean and Black seas. The application of a genome-wide approach based on single nucleotide polymorphism (SNP) markers, a wide sampling design and a good sample size, allowed to overcome the previous limitations providing the first reliable biogeographic data concerning this species throughout its native range.

Connectivity data obtained for these 2 species, together with the overall results from all the CoCoNET species, will be taken into account to improve the effectiveness of environmental and management policies in the 2 basins, helping to determine whether existing MPAs are effective in sustaining ecological networks and suggesting how to design better protection systems based on an efficient connection between the protected areas.

Following contents

- **CHAPTER 2 – Seascape genomics of the sea urchin *Paracentrotus lividus* in the Central Mediterranean Sea**

This Chapter reports the results of the sea urchin. The Chapter includes my sea urchin's paper and its Supporting Information, recently published on *Ecology & Evolution* as:

Paterno M, Schiavina M, Aglieri G, Ben Souissi J, Boscari E, Casagrandi R, Chassanite A, Chiantore M, Congiu L, Guarnieri G, Kruschel C, Macic V, Marino IAMM, papetti C, Patarnello T, Zane L, Melià P (2017) Population genomics meet Lagrangian simulations: oceanographic patterns and long larval duration ensure connectivity among *Paracentrotus lividus* populations in the Adriatic and Ionian seas. *Ecology & Evolution*, 00, 1–17, doi: 10.1002/ece3.2844.

- **CHAPTER 3 – Genomic population structure of the mussel *Mytilus galloprovincialis* in the Mediterranean and Black seas**

This Chapter reports the results of the Mediterranean mussel. The results will be submitted soon to the journal *Diversity & Distributions* as:

Paterno M, Ben Souissi J, Boscari E, Chassanite A, Congiu L, Guarnieri G, Kruschel C, Macic V, Marino IAM, Micu D, Voutsinas M, Papetti C, Zane L. A genome-wide approach to the global phylogeography of the mussel *Mytilus galloprovincialis* in the Mediterranean and Black Seas.

- **CHAPTER 4 – Seascape genetics of the marbled crab *Pachygrapsus marmoratus* in the Central Mediterranean**

This Chapter reports the title page, the abstract and an extended summary of a further seascape genetics paper in which I was involved. I collaborated mainly in data analysis ad interpretation. Since I'm not the first author, I decided to not report the entire manuscript. The paper is currently under revision to journal *Molecular Ecology* as:

Marino IAM, Schiavina M, Aglieri G, Bevilacqua S, Boscari E, Congiu L, Faggion S, Kruschel C, Papetti C, Patarnello T, **Paterno M**, Voutsinas E, Zane L, Melià P. Combining genetic data and Lagrangian simulations to assess connectivity patterns of the marbled crab *Pachygrapsus marmoratus* in the Adriatic and Ionian seas.

- **CHAPTER 5 – Conclusion**

This Chapter reports a comparison of the results obtained for the sea urchin, the Mediterranean Sea, the marble crab and others CoCoNET species, in the light of the implications for the MPAs management in the Mediterranean and Black seas.

References

Airoldi L, Beck MW (2007) Loss, status and trends for coastal marine habitats of Europe. *Oceanography and Marine Biology: An Annual Review*, 45, 345–405.

Almany GR, Berumen ML, Thorrold SR, Planes S, Jones GP (2007) Local replenishment of coral reef fish populations in a marine reserve. *Science*, 316: 742–744.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Review Genetics*, 17, 81–92.

Bader JM (1998) Measuring Genetic Variability in Natural Populations by Allozyme Electrophoresis. Pages 25-42, in *Tested studies for laboratory teaching*, Volume 19 (Karcher SJ, Editor).

Baird NA, Etter PD, Atwood TS, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3, e3376.

Ballard JWO, Whitlock MC (2004) The Incomplete Natural History of Mitochondria. *Molecular Ecology*, 13, 729–744.

Baxter SW, Davey JW, Johnston JS et al. (2011) Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS ONE*, 6, e19315.

Bohonak AJ (1999) Dispersal, gene flow, and population structure. *The Quarterly Review of Biology*, 74, 21–45.

Bonneaud C, Balenger S, Russell AF, Zhang J, Hill G, Edwards SV (2011) Rapid evolution of disease resistance is accompanied by functional changes in gene expression in a wild bird. *Proceedings of the National Academy of Sciences USA*, 108, 7866–7871.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32, 314–331.

Boudouresque et al. 2005

Bowen BW, Bass AL, Muss A, Carlin J, Robertson DR (2006) Phylogeography of Two Atlantic Squirrelfishes (Family Holocentridae): Exploring Links between Pelagic Larval Duration and Population Connectivity. *Marine Biology*, 149, 899–913.

Broquet T, Petit EJ (2009) Molecular estimation of Dispersal for Ecology and Population Genetics. *Annual Review of Ecology, Evolution, and Systematics*, 40, 193–216.

Campana SE (1999) Chemistry and composition of fish otoliths: pathways, mechanisms and applications. *Marine Ecology Progress Series*, 188, 263–297.

Carson HS, Cook GS, Lopez-Duarte PC, Levin LA (2011) Evaluating the importance of demographic connectivity in a marine metapopulation. *Ecology*, 92, 1972–1984.

Claudet J, Osenberg CW, Benedetti-Cecchi L et al. (2008) Marine reserves: size and age do matter. *Ecology Letters*, 11, 481–489.

Claudet J, Fraschetti S (2010) Human impacts on marine habitats: a regional meta-analysis in the Mediterranean Sea. *Biological Conservation*, 143, 2195–2206.

Cowen RK, Paris CB, Srinivasan A (2006) Scaling of Connectivity in Marine Populations. *Science*, 311, 522–527.

Cowen RK, Gawarkiewicz G, Pineda J, Thorrold SR, Werner FE (2007) Population connectivity in marine systems: An overview. *Oceanography*, 20, 14–21.

Cowen RK, Sponaugle S (2009) Larval Dispersal and Marine Population Connectivity. Annual Review of *Marine Science*, 1, 443–466.

Crooks KR, Sanjayan M (2006) *Connectivity conservation*. Cambridge University Press, Cambridge, UK.

D'Aloia CC, Bogdanowicz SM, Francis RK, Majoris JE, Harrison RG, Buston PM (2015) Patterns, Causes, and Consequences of Marine Larval Dispersal. *Proceedings of the National Academy of Sciences*, 112, 13940–13945.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12, 499–510.

Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, 9, 416–423.

Di Franco A, Coppini G, Pujolar JM, et al. (2012) Assessing Dispersal Patterns of Fish Propagules from an Effective Mediterranean Marine Protected Area. *PLoS ONE*, 7, e52108

Doherty PJ, Planes S, Mather P (1995) Gene flow and larval duration in seven species of fish from the Great Barrier Reef. *Ecology*, 76, 2373–2391.

Dupont L, Ellien C, Viard F (2007) Limits to gene flow in the slipper limpet *Crepidula fornicata* as revealed by microsatellite data and a larval dispersal model. *Marine Ecology Progress Series*, 349, 125–138.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6:e19379.

Emlet RB, McEdward LR, Strathmann RR (1987) Echinoderm larval ecology viewed from the egg. *Echino- derm Studies*, 2, 55–136.

Etter PD, Preston JL, Bassham S, Cresko WA, Johnson ES (2011) Local De Novo Assembly of RAD Paired-End Contigs Using Short Sequencing Reads. *PLoS ONE*, 6, e18561.

Fraschetti S, Guarnieri G, Bevilacqua S et al. (2011) Conservation of Mediterranean habitats and biodiversity countdowns: what information do we really need? *Aquatic conservation: marine and freshwater ecosystems*, 21, 299–306.

Fenberg PB, Caselle JE, Claudet J et al. (2012) The Science of European Marine Reserves: Status, Efficacy, and Future Needs. *Marine Policy*, 36, 1012–21.

Gabrié C, Lagabrielle E, Bissery C et al. (2012) The Status of Marine Protected Areas in the Mediterranean Sea. MedPAN & RAC/SPA. Ed: MedPAN Collection.

Gaines SD, White C, Carr MH, Palumbi SR (2010), Designing marine reserve networks for both conservation and fisheries management. *PNAS*, 43, 18286–18293.

Galindo HM, Olson DB, Palumbi SR (2006) Seascape genetics: a coupled oceanographic–genetic model predicts population structure of Caribbean corals. *Current Biology*, 16, 1622–1626.

Gilg MR, Hilbish TJ (2003) The Geography of Marine Larval Dispersal: Coupling Genetics with Fine-scale Physical Oceanography. *Ecology*, 84, 2989–2998.

Grorud-Colvert K, Claudet J, Tissot BN et al. (2014) Marine protected area networks: assessing whether the whole is greater than the sum of its parts. *PLoS ONE*, 9, e102298.

Guidetti P, Sala E (2007). Community-wide effects of marine reserves in the Mediterranean Sea. *Marine Ecology Progress Series*, 335, 43–56.

Halpern BS, Selkoe KA, Micheli F, Kappel CV (2007) Evaluating and ranking the vulnerability of global marine ecosystems to anthropogenic threats. *Conservation Biology*, 21, 1301–1315.

Hansen MM, Hemmer-Hansen J (2007) Landscape genetics goes to sea. *Journal of Biology*, 6, 6.

Hastings A, Harrison S (1994) Metapopulation Dynamics and Genetics. *Annual Review of Ecology and Systematics*, 25, 167–188.

Hellberg ME, Burton RS, Neigel JE, Palumbi SR (2002) Genetic Assessment of Connectivity among Marine Populations. *Bulletin of Marine Science*, 70, 273–290.

Hohenlohe, P.A. (2004). Limits to gene flow in marine animals with planktonic larvae: models of *Littorina* species around Point Conception, California. *Biological Journal of the Linnena Society of London*, 82, 169–187.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson E, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, 6, e1000862.

Hulme (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46, 10–18.

Jablonski D (1986) Larval ecology and macroevolution in marine invertebrates. *Bulletin of Marine Sciences*, 39, 565–587.

Jackson JBC, Kirby MX, Berger WH et al. (2001) Historical overfishing and the recent collapse of coastal ecosystems. *Science*, 293, 629–637.

Jarne P, Lagoda PJ (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, 11, 424–429.

Kool JT, Moilanen A, Treml EA (2013) Population connectivity: recent advances and new perspectives. *Landscape Ecology*, 28, 165–185.

Kenchington EL, Patwary MU, Zouros E, Bird CJ (2006) Genetic differentiation in relation to marine landscape in a broadcast-spawning bivalve mollusc (*Placopecten magellanicus*). *Molecular Ecology*, 15, 1781–1796.

Kinlan BP, Gaines SD, Lester SE (2005) Propagule dispersal and the scales of marine community process. *Diversity and Distribution*, 11, 139–148.

Lee T, Clarke ME, Williams E, Szmant AF, Berger T (1994) Evolution of the Tortugas Gyre and its influence on recruitment in the Florida Keys. *Bulletin of Marine Science*, 54:621–646.

Lejeusne C, Chevaldonne P, Pergent-Martini C, Boudouresque CF, Perez T (2010) Climate change effects on a miniature ocean: the highly diverse, highly impacted Mediterranean Sea. *Trends in Ecology & Evolution*, 25, 250–260.

Lester SE, Ruttenberg BI, Gaines SD, Kinlan BP (2007) The relationship between dispersal ability and geographic range size. *Ecology Letters*, 10, 745–758.

Levin LA (2006) Recent progress in understanding larval dispersal: new directions and digressions. Integrative & Comparative Biology, 46, 282–297.

Lòpez-Duarte P, Carson H, Cook G et al. (2012) What Controls Connectivity? An Empirical, Multi-Species Approach. *Integrative and Comparative Biology*, 52, 511–524.

Lotze HK, Lenihan HS, Bourque BJ et al. (2006) Depletion, degradation, and recovery potential of estuaries and coastal seas. *Science*, 312, 1806–1809.

Lowe W, Allendorf F (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, 19, 3038–3051.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Review Genetics*, 4, 981–994.

Marino IAM, Schiavina M, Aglieri G et al. (in review, *Molecular Ecology*) Assessing Connectivity Patterns of the Marbled Crab *Pachygrapsus marmoratus* in the Adriatic and Ionian Seas Combining Genetic Data and Lagrangian Simulations.

Marshall JJT, Halford SE (2010) The Type IIB restriction endonucleases. *Biochemical Society Transactions*, 38, 410–416.

Melià P, Schiavina M, Rossetto M, Gatto M, Fraschetti S, Casagrandi R (2016) Looking for Hotspots of Marine Metacommunity Connectivity: a Methodological Framework. *Scientific Reports*, 6, 23705.

Mileikovsky SA (1971) Types of Larval Development in Marine Bottom Invertebrates, Their Distribution and Ecological Significance: A Re-Evaluation. *Marine Biology*, 10, 193–213.

Moritz C (1994) Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution*, 9, 373–375.

Munday P, Leis J, Lough J, Paris C, Kingsford M, Berumen M, Lambrechts J (2009) Climate change and coral reef connectivity. *Coral Reefs*, 28, 379–395.

Narum SR, Campbell NR, Kozfkay CC, Meyer KA (2010) Adaptation of redband trout in desert and montane environments. *Molecular Ecology*, 19, 4622–4637.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22, 2841–2847.

Palumbi SR (2003) Population Genetics, Demographic Connectivity, and the Design of Reserves. *Ecological Applications*, 13, 146–158.

Paterno M, Schiavina M, Aglieri G et al. (2017) Population genomics meet Lagrangian simulations: oceanographic patterns and long larval duration ensure connectivity among *Paracentrotus lividus* populations in the Adriatic and Ionian seas. *Ecology & Evolution*, 00, 1–17.

Pauletto M, Carraro L, Babbucci M, Lucchini R, Bargelloni L, Cardazzo B (2016) Extending RAD tag analysis to microbial ecology: a comparison between MultiLocus Sequence Typing and 2b-RAD to investigate *Listeria monocytogenes* genetic structure. *Molecular Ecology Resources*, 16, 823–35.

Pechenik JA (1990) Delayed metamorphosis by larvae of benthic marine invertebrates: Does it occur? Is there a price to pay? *Ophelia*, 32, 63–94.

Pecoraro C, Babbucci M, Villamor A et al. (2016) Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*). *Marine Genomics*, 25, 43–8.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7:e37135.

Puckett BJ, Eggleston DB, Kerr PC, Luettich RA Jr (2014) Larval Dispersal and Population Connectivity among a Network of Marine Reserves. *Fisheries Oceanography*, 23, 342–361.

Pujolar JM, Schiavina M, Di Franco A et al. (2013) Understanding the effectiveness of marine protected areas using genetic connectivity patterns and Lagrangian simulations. *Diversity and Distributions*, 19, 1531–1542.

Ruzzante DE, Taggart CT, Cook D (1998) A nuclear DNA basis for shelf- and bank-scale population structure in northwest Atlantic cod (Gadus morhua): Labrador to Georges Bank. *Molecular Ecology*, 7, 1663–1680.

Sale P, Cowen R, Danilowicz et al. (2005). Critical science gaps impede use of no-take fishery reserves. *Trends in Ecology & Evolution*, 20, 74–80.

Schiavina M, Marino IAM, Zane L, Melià P (2014) Matching oceanography and genetics at the basin scale. Seascape connectivity of the Mediterranean shore crab in the Adriatic Sea. *Molecular Ecology*, 23, 5496–5507.

Seber GAF (1982) *The estimation of animal abundance and related parameters*, 2nd ed. MacMillan, New York.

Selkoe KA, D'Aloia CC, Crandall ED et al. (2016) A decade of seascape genetics: contributions to basic and applied marine connectivity. *Marine Ecology Progress Series*, 554, 1–19.

Selkoe KA, Toonen RJ (2011) Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series*, 436, 291−305.

Selkoe KA, Watson JR, White C et al. (2010) Taking the Chaos out of Genetic Patchiness: Seascape Genetics Reveals Ecological and Oceanographic Drivers of Genetic Patterns in Three Temperate Reef Species. *Molecular Ecology*, 19, 3708–3726.

Selkoe KA, Henzler CM, Gaines SD (2008) Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries*, 9, 363–377.

Shanks AL, Grantham BA, Carr MH (2003) Propagule dispersal distance and the size and spacing of marine reserves. *Ecological Applications*, 13, S159–S169.

Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, Normandeau É, Laroche J, Larose S, Jean M (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS ONE*, 8:e54603.

Sponaugle S, Cowen RK, Shanks A, Morgan SG et al. (2002) Predicting self-recruitment in marine populations: biophysical correlates and mechanisms. *Bulletin of Marine Sciences*, 703, 41–375.

Taylor MS, Hellberg ME (2003) Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. *Science*, 299, 107–109.

Thorrold SR, Jones GP, Hellberg ME et al. (2002) Quantifying Larval Retention and Connectivity in Marine Populations with Artificial and Natural Markers. *Bulletin of Marine Science*, 70, 291–308.

Thorrold SR, Zacherl DC, Levin LA (2007) Population connectivity and larval dispersal using geochemical signatures in calcified structures. *Oceanography*, 20, 80–89.

Todd CD (1998) Larval Supply and Recruitment of Benthic Invertebrates: Do Larvae Always Disperse as Much as We Believe? *Hydrobiologia*, 375, 1–21.

Toonen RJ, Puritz JB, Forsman ZH, Whitney JL, Fernandez-Silva I, Andrews KR, Bird CE (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ*, 1, e203.

Vos P, Hogers R, Bleeker M et al. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res*, 23, 4407–4414.

Wabnitz CCC, Andrefouet S, Muller-Karger FE (2010) Measuring progress toward global marine conservation targets. *Frontiers in Ecology and the Environment*, 8, 124–129.

Wang S, Meyer E, McKay JK, Matz MV (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods*, 9, 808–810.

Waples RS (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, 89, 438–450.

Webster MS, Marra PP, Haig SM et al. (2002) Links between worlds: unraveling migratory connectivity. *Trends in Ecology & Evolution*, 17, 76–83.

Weersing K, Toonen RJ (2009) Population Genetics, Larval Dispersal, and Connectivity in Marine Systems. *Marine Ecology Progress Series*, 393, 1–12.

White C, Selkoe KA, Watson J, Siegel DA, Zacherl DC, Toonen RJ (2010) Ocean Currents Help Explain Population Genetic Structure. *Proceedings of the Royal Society of London*, Series B, 277, 1685–1694.

Wood LJ, Fish L, Laughren J, Pauly D (2008) Assessing progress towards global marine protection targets: shortfalls in information and action. *Oryx*, 42, 340–351.

Woodson CB, McManus MA (2007) Foraging behavior can influence dispersal of marine organisms. *Limnology and Oceanography*, 52, 2701–2709.

Worm B, Barbier EB, Beaumont N et al. (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science*, 314, 787–790.

Wright S (1931) Evolution in Mendelian Populations. *Genetics*, 16, 97–159.

Xu T, Sun J, Lv J et al. (2016) Genome-wide discovery of single nucleotide polymorphisms (SNPs) and single nucleotide variants (SNVs) in deep-sea mussels: Potential use in population genomics and cross-species application. *Deep-Sea Research II*, 137, 318–326.

CHAPTER 2

Seascape genomics of the sea urchin *Paracentrotus lividus* in the Central Mediterranean Sea

Published as:

Paterno M, Schiavina M, Aglieri G et al. (2017) **Population genomics meet Lagrangian simulations: oceanographic patterns and long larval duration ensure connectivity among *Paracentrotus lividus* populations in the Adriatic and Ionian seas.** *Ecology & Evolution*, 00, 1–17, doi: 10.1002/ece3.2844

Received: 2 November 2016 | Revised: 20 January 2017 | Accepted: 28 January 2017

DOI: 10.1002/ece3.2844

ORIGINAL RESEARCH

WILEY Ecology and Evolution Open Access

Population genomics meet Lagrangian simulations: Oceanographic patterns and long larval duration ensure connectivity among *Paracentrotus lividus* populations in the Adriatic and Ionian seas

Marta Paterno[1,2] | Marcello Schiavina[2,3] | Giorgio Aglieri[2,4] | Jamila Ben Souissi[5] | Elisa Boscari[1,2] | Renato Casagrandi[2,3] | Aurore Chassanite[6] | Mariachiara Chiantore[2,7] | Leonardo Congiu[1,2] | Giuseppe Guarnieri[2,4] | Claudia Kruschel[8] | Vesna Macic[9] | Ilaria A. M. Marino[1,2] | Chiara Papetti[1,2] | Tomaso Patarnello[2,10] | Lorenzo Zane[1,2] | Paco Melià[2,3]

[1]Department of Biology, University of Padova, Padova, Italy

[2]Consorzio Nazionale Interuniversitario per le Scienze del Mare (CoNISMa), Roma, Italy

[3]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy

[4]Department of Biological and Environmental Sciences and Technologies, University of Salento, Lecce, Italy

[5]Institut National Agronomique de Tunisie (INAT), Tunis, Tunisia

[6]USR 3278 CNRS-EPHE, CRIOBE, Université de Perpignan Via Dominitia, Perpignan Cedex, France

[7]Department for Earth, Environment and Life Sciences (DiSTAV), University of Genoa, Genoa, Italy

[8]University of Zadar, Zadar, Croatia

[9]Institute of Marine Biology Kotor (IBMK), Kotor, Montenegro

[10]Department of Comparative Biomedicine and Food Science, University of Padova, Legnaro, Padova, Italy

Correspondence
Lorenzo Zane, Department of Biology, University of Padova, Padova, Italy.
Email: lorenzo.zane@unipd.it

Abstract

Connectivity between populations influences both their dynamics and the genetic structuring of species. In this study, we explored connectivity patterns of a marine species with long-distance dispersal, the edible common sea urchin *Paracentrotus lividus*, focusing mainly on the Adriatic–Ionian basins (Central Mediterranean). We applied a multidisciplinary approach integrating population genomics, based on 1,122 single nucleotide polymorphisms (SNPs) obtained from 2b-RAD in 275 samples, with Lagrangian simulations performed with a biophysical model of larval dispersal. We detected genetic homogeneity among eight population samples collected in the focal Adriatic–Ionian area, whereas weak but significant differentiation was found with respect to two samples from the Western Mediterranean (France and Tunisia). This result was not affected by the few putative outlier loci identified in our dataset. Lagrangian simulations found a significant potential for larval exchange among the eight Adriatic–Ionian locations, supporting the hypothesis of connectivity of *P. lividus* populations in this area. A peculiar pattern emerged from the comparison of our results with those obtained from published *P. lividus* cytochrome b (cytb) sequences, the latter revealing genetic differentiation in the same geographic area despite a smaller sample size and a lower power to detect differences. The comparison with studies conducted using nuclear markers on other species with similar pelagic larval durations in the same Adriatic–Ionian locations indicates species-specific differences in genetic connectivity patterns and warns against generalizing single-species results to the entire community of rocky shore habitats.

KEYWORDS

2b-RAD, biophysical models, population genomics, sea urchin, seascape genetics, SNPs

1 | INTRODUCTION

Population connectivity plays a key role in evolutionary and ecological processes that shape population dynamics and genetic structuring of species (Cowen, Paris, & Srinivasan, 2006; Puckett, Eggleston, Kerr, & Luettich, 2014). The ability to define the spatial scale at which population connectivity occurs is fundamental to enhancing our knowledge of dynamics and persistence of marine metapopulations and improving the success of biological resources management (Palumbi, 2003).

Most marine organisms (approximately 70% of marine invertebrates; Mileikovsky, 1971) display a bipartite life cycle characterized by a pelagic larval phase and sedentary adults. In these species, larval dispersal is expected to be the main process driving the exchange of individuals across seascapes and, consequently, ensuring population connectivity (D'Aloia et al., 2015; Gilg & Hilbish, 2003).

The small size of larvae, as well as the rapid dilution with distance and time from their natal origin, makes direct tracking of larval dispersal impractical (Cowen & Sponaugle, 2009). The study of genetic isolation patterns provides an indirect, often powerful, source of information for the inference of connectivity among populations (Hellberg, Burton, Neigel, & Palumbi, 2002; Palumbi, 2003; Thorrold et al., 2002). A review of 300 population genetic studies by Weersing and Toonen (2009) reported a weak correlation between average pelagic larval duration (PLD) and genetic differentiation (F_{ST}) in marine taxa. In several cases, strong genetic partitioning was found despite a long PLD, indicating that marine populations can be less connected than expected. Limited connectivity between populations, even at small spatial scales (Taylor & Hellberg, 2003), may be explained by a number of mechanisms related to the biology of each species, such as larval behavior (Todd, 1998) and specific life-history traits (Bowen, Bass, Muss, Carlin, & Robertson, 2006), or to ecological and environmental factors (Cowen & Sponaugle, 2009). Due to the complexity of the marine environment, Euclidean distance between locations is often a poor predictor of realized dispersal distance, because water circulation patterns and mesoscale features such as eddies and gyres can determine retention and/or diffusion of pelagic larvae at very different scales (Ruzzante, Taggart, & Cook, 1998; White et al., 2010).

Seascape genetics combines genetics and oceanography to investigate marine population connectivity by testing the role of environmental drivers in shaping the spatial genetic structure of species (Hansen & Hemmer-Hansen, 2007; Storfer et al., 2006; White et al., 2010), thus allowing the comparison of the two approaches (Galindo, Olson, & Palumbi, 2006; Galindo et al., 2010; Selkoe et al., 2010).

In our study, we applied a seascape genetics approach to explore connectivity patterns of a species with a long PLD, the edible common sea urchin *Paracentrotus lividus* (Lamarck, 1816; Figure 1), with special focus on populations in the Adriatic and Ionian seas. This geographic area was selected in the FP7 CoCoNET project (http://www.coconet-fp7.eu/) to investigate connectivity patterns of different species and to understand the scale at which marine protected areas (MPAs) can work as an effective network. We combined a population genomics approach with Lagrangian simulations produced by a biophysical model coupling oceanographic reanalyses (used as drivers for the movement of the planktonic stage) with information on life-history traits affecting species dispersal. A set of genome-wide single nucleotide polymorphisms (SNPs) was obtained using 2b-RAD (Wang, Meyer, McKay, & Matz, 2012), a simple and cost-effective restriction-site-associated DNA (RAD) sequencing method for genome reduction (Baird et al., 2008) aimed at the simultaneous identification and genotyping of thousands of SNPs evenly spread across the genome. This next-generation sequencing approach provides a much higher resolution compared to those based on traditional markers, such as mitochondrial DNA sequencing or microsatellite genotyping. In addition, yielding a high number of loci scattered across the genome allows for the investigation into the presence of nonneutral processes of differentiation (Stapley et al., 2010) through the detection of outliers (Excoffier, Hofer, & Foll, 2009).

FIGURE 1 *Paracentrotus lividus* (Kroh & Mooi, 2017)

Echinoderms play a key role in structuring marine ecosystems. The sea urchin *P. lividus* is a keystone species in benthic communities of Mediterranean infralittoral rocky shores. Its grazing activity is considered one of the principal controlling factors of macroalgal community structure (Boudouresque & Verlaque, 2001; Privitera, Noli, Falugi, & Chiantore, 2011). This essential ecological role is threatened, because the species' gonads are considered a culinary delicacy, and therefore, sea urchins are heavily exploited at regional scales (Barnes & Crook, 2001). *P. lividus* is found in the sublittoral zone down to 20 m, throughout a distribution range that includes the Mediterranean Sea and the North-East Atlantic Ocean from Ireland to the coasts of Morocco (Boudouresque & Verlaque, 2001). The larval stage (echinopluteus) is, on average, 30 days long (Fenaux, Cellario, & Etienne, 1985; Pedrotti, 1993).

The genetic structure of this species has been recently explored throughout its distribution range using mitochondrial and nuclear DNA sequencing (Calderón, Giribet, & Turon, 2008; Duran, Palacín, Becerro, Turon, & Giribet, 2004; Maltagliati, Di Giuseppe, Barbieri, Castelli, & Dini, 2010; Penant, Didier, Feral, & Chenuil, 2013). These studies detected the presence of two major genetic discontinuities: one between the Atlantic Ocean and the Mediterranean Sea, explained by the Almería-Oran hydrological front (Calderón et al., 2008; Duran

TABLE 1 Sampling information of *Paracentrotus lividus* population samples examined in this study. For each population sample, sampling information about area, nation, sampling location, acronym, coordinates, date, and number of individuals *N* (processed/analyzed) are reported

Area	Nation	Sampling location	Acronym	Coordinates	Date	*N*
Ionian Sea	Greece	Othonoi Island	OTH	39.793289N 19.935636E	July 2013	31/31
Ionian Sea	Albania	Karaburun Peninsula	KAP	40.392800N 19.324967E	June 2013	30/30
Adriatic Sea	Montenegro	Boka Kotorska	BOK	42.387533N 18.569633E	June 2013	26/25
Adriatic Sea	Croatia	Kornati Islands	KOR	43.792250N 15.281483E	June 2013	26/26
Adriatic Sea	Italy	Tremiti Islands	TRE	42.138583N 15.523950E	April 2013	29/28
Adriatic Sea	Italy	Torre Guaceto	TOG	40.716650N 17.800050E	May 2013	31/31
Adriatic–Ionian Sea	Italy	Otranto	OTR	40.109233N 18.519217E	May 2013	30/30
Ionian Sea	Italy	Porto Cesareo	POC	40.195250N 17.917950E	May 2013	30/29
Western Med. Sea	France	Banyuls	FRN	42.482290N 3.1374160E	October 2014	26/26
Western Med. Sea	Tunisia	Haouaria	TUN	37.050440N 10.967000E	January 2014	16/16

et al., 2004; Maltagliati et al., 2010), and the other one between the Adriatic Sea and the rest of the Mediterranean Sea (Maltagliati et al., 2010). By adding new mitochondrial and nuclear sequences and by reanalyzing published datasets, Penant et al. (2013) recently detected a signal of genetic differentiation at a smaller geographic scale, including differentiation within the Adriatic–Ionian seas.

In this context, thanks to their high frequency and genome-wide coverage, SNP markers are expected to provide a more reliable and detailed picture of the genetic differentiation of *P. lividus*. Specifically, our study aims to: (1) evaluate the presence of large-scale genetic structuring by comparing eight population samples from the Central Mediterranean (Adriatic and Ionian seas) with two population samples from the Western Mediterranean (i.e., France and Tunisia); (2) investigate the presence of genetic differentiation at a smaller scale by comparing samples within and between the Adriatic and Ionian basins; (3) estimate potential larval connectivity and retention via Lagrangian simulations based on a biophysical model developed for the Adriatic and Ionian basins; (4) compare the patterns obtained in (3) with the realized connectivity estimated via genetic analyses, in order to provide integrated and reliable data concerning the dispersal scale of *P. lividus*; (5) obtain useful information for the planning and management of MPAs across this geographic area by comparison with other target species.

2 | MATERIALS AND METHODS

2.1 | Collection of samples, species confirmation, and total DNA extraction

A total of 275 individuals were collected at eight sampling locations in the Central Mediterranean (Adriatic–Ionian seas, FAO subarea 37.2; FAO 2004) and two locations in the Western Mediterranean (France and Tunisia, FAO subarea 37.1; FAO 2004) between April 2013 and October 2014 (Table 1 and Figure 2). Sampling locations in the

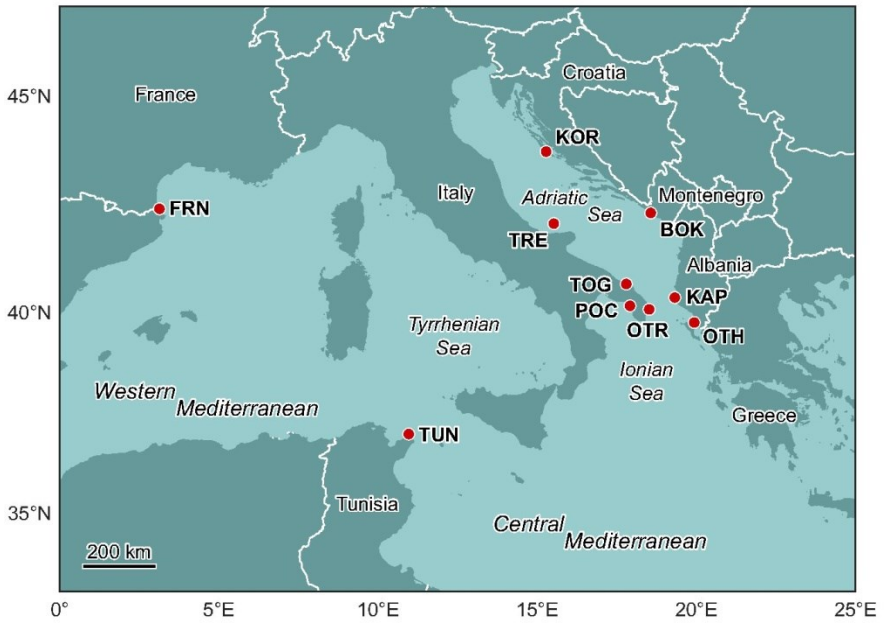


FIGURE 2 Sampling sites in the Central (Adriatic–Ionian seas) and the Western Mediterranean Sea (FAO subareas 37.1 and 37.2; FAO 2004). See Table 1 for location acronyms

Adriatic–Ionian seas were at nearest neighbors' distances between 70 and 300 km. The two additional population samples from France and Tunisia were included for comparison at a larger scale. At each location, sea urchins were collected by scuba divers. For each individual, 5–8 podia were removed and preserved in 95–96% ethanol until DNA extraction. Total genomic DNA (gDNA) was extracted using the Eurogold Tissue DNA Mini Kit (EuroClone), and its integrity was assessed by visualization on a 1% agarose gel stained with GelRed (BIOTIUM, GelRed™ Nucleic Acid Stain, 10,000× in Water). Concentrations and purity ratios of the samples (260/280 nm and 260/230 nm) were obtained by means of a NanoDrop UV–Vis spectrophotometer. When needed, samples were concentrated by precipitation with isopropanol (Sambrook, Fritsch, & Maniatis, 1990). Species identification was confirmed by PCR amplification and sequencing of about 1,000 bp (base pairs) of the cytochrome b gene (cytb) in at least 10 individuals from each population sample (N = 119). To this end, amplification and sequencing primers from Maltagliati et al. (2010) were used with the following PCR conditions: a total volume of 20 μl containing 2 mmol/L $MgCl_2$, 0.2 μmol/L of each primer (Invitrogen), 250 μmol/L of each dNTPs, 1× Reaction Buffer (RBC Bioscience), 0.05 U/μl of *Taq* DNA polymerase (RBC Bioscience), 20 ng of DNA, and an amplification profile consisting of six touchdown cycles of 60 s at 94°C, 30 s at 62°C with a decrease of 0.7°C per cycle and 120 s at 72°C, followed by 24 cycles of 60 s at 94°C, 30 s at 58°C, and 120 s at 72°C, and by 5 min at 72°C for final extension. PCR products were purified using QIAquick PCR Purification Kit (QIAGEN) following the instructions of the supplier and sequenced by an external service (http://www.bmr-genomics.it/). Sequences were compared to NCBI database (blastn) to confirm identification of samples as *Paracentrotus lividus*.

2.2 | Construction of 2b-RAD libraries

The 2b-RAD technique (Wang et al., 2012) was tested on six samples and optimized for processing *P. lividus* gDNA (2b-RAD oligonucleotide sequences used for Illumina sequencing are reported in Table S1). Following this preliminary test, about 400 ng of high-quality RNA-free gDNA from each individual (N = 275) was cleaved with 2 U of the type 2b restriction endonucleases *Csp*CI (New England BioLabs) overnight at 37°C, producing a population of fragments of uniform length (35 bp) with protruding ends (Marshall & Halford, 2010). For each sample, 1 μl of digested DNA was loaded on a 1% agarose gel stained with GelRed™ alongside a comparable amount of intact genomic DNA from the same sample to check the performance of the restriction reaction. The digested products were ligated to partially double-stranded adaptors with compatible and fully degenerated overhangs in a 25-μl total volume reaction consisting of 0.4 μmol/L of each adaptor, 0.2 mmol/L ATP, and 40 U/μl T4 DNA ligase (SibEnzyme). The 2b-RAD tags were amplified for a few cycles (16) using two pairs of primers to amplify and to introduce sample-specific barcodes (7 bp) and the annealing sites for Illumina next-generation sequencing. Barcodes were designed by Barcode Generator (available at UC Davis Web site: http://comailab.genomecenter.ucdavis.edu/index.php/Barcode_generator). Amplification PCR consisted of 10 μl of ligated DNA, 0.2 μmol/L of each primer 2b-RAD amp, 0.5 μmol/L of each primer F and barcoded primer R, 0.3 mmol/L dNTPs, 1× Phusion HF buffer, and 0.02 U/μl Taq Phusion high-fidelity polymerase (New England BioLabs) in a total volume of 50 μl. Amplification products were run on 1.8% agarose gel stained with GelRed™ to check the quality of reactions. PCR products (1–10 μl) of different individuals were pooled according to the concentration of the target band (about 170 bp, including the restriction fragment of interest and the adaptors/barcodes). The amount of the PCR product from each individual was estimated by combining the concentration values obtained from NanoDrop UV–Vis spectrophotometer and the target band intensity obtained from agarose gel using ImageJ (Schneider, Rasband, & Eliceiri, 2012). A total of three pools (about 90 barcoded samples per pool) were assembled. The target band of about 170 bp was purified by removing contaminating fragments (high molecular weight fragments and primer–dimers) in two consecutive steps. First, each pool was run on 1% agarose gel, and the target band was excised and eluted in distilled water overnight; second, a further purification of each pool was performed using magnetic beads (SPRIselect, BECKMAN COULTER) according to the solid-phase reversible immobilization (SPRI) method (DeAngelis, Wang, & Hawkins, 1995).

2.3 | 2b-RAD tag sequencing

Each pool was sequenced on Illumina HiSeq platforms with a single-end SR50 High Output mode by UC Davis Genome Center (CA, USA) or Genomix4Life S.r.l. (Baronissi, SA, Italy), which also performed data demultiplexing and quality filtering. Each pool was sequenced twice: following sequencing results of the first run, the relative amount of the target 170 bp band of each individual was precisely estimated based on the number of reads obtained for each sample. A new pool was then assembled for each group of individuals, adjusting volumes to obtain the final normalization. The new pool was purified and sequenced as before, and the reads of each individual from the first and second Illumina sequencing run were merged and analyzed together. This procedure allowed for an increase in the depth coverage and ensured obtaining an equal and comparable number of reads from each analyzed sample.

2.4 | *De novo* analysis: in silico identification of loci and genotyping

The quality of raw demultiplexed reads was checked with FastQC software (available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/); then, custom-made scripts were used to filter reads for the presence of *Csp*CI recognition sites and to trim adaptors, obtaining sequences of uniform length (32 bp). The trimmed, high-quality reads (mean quality score per base > 37) were used for subsequent analysis. In silico assembly of loci and genotyping was performed using STACKS software (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) employing the "denovo_map.pl" pipeline. A technical replicate of one sample was used to estimate error rates

and to optimize *de novo* assembly parameters of 2b-RAD data. The parameters were set to: minimum stack depth (*m*) of 15; number of mismatches allowed between stacks to build a locus in an individual (*M*) equal to 4; maximum distance between loci from distinct individuals to be merged in the population catalog (*n*) equal to 4; error rate to call SNP (*bound*) between 0 and 0.1; significance level required to call a heterozygote or homozygote (*alpha*) equal to 0.05. The STACKS module *Populations* was used to generate output in GENEPOP format for further downstream analysis. As the minimum coverage for each allele per locus was set to 15 (parameter *m*, see above), no further filtering by sequencing depth was applied. CREATE software (Coombs, Letcher, & Nislow, 2008) was used for conversion of the GENEPOP input file into input formats for different genetic analysis programs. The final dataset included all polymorphic loci present in at least 80% of the individuals, and it was characterized by the presence of 1–3 SNPs and 2–6 alleles. When multiple SNPs were found at one locus, only the SNP with the highest expected heterozygosity (proxy for polymorphic content, Phillips, 2005) across the whole dataset was retained. The threshold of missing loci per individual was set to 30%.

2.5 | SNPs validation by Sanger sequencing

To evaluate the accuracy of the 2b-RAD protocol and to verify the existence and polymorphism level of candidate SNPs, a preliminary STACKS run of 136 individuals was used to select a reduced number of loci for validation by a PCR-based approach and Sanger sequencing. An available dataset of more than 140,000 EST-linked sequences of *P. lividus* was downloaded from the NCBI database (four libraries; accession numbers: MPMGp1171–1174) and used to create a local database. Consensus sequences of a subset of 2b-RAD loci with different polymorphism levels were used as query for a local blastn against the EST-linked sequences database. Consensus sequences of each locus and its matching EST-linked sequences were aligned using CLUSTALW Omega tool (available online at http://www.ebi.ac.uk). Primers for Sanger sequencing were designed on a selection of 10 loci using sequence information of alignments and tested on a small number of individuals with a known genotype. PCR conditions were optimized depending on annealing temperature of each pair of primers and on the expected amplicon length (primer sequences used for validated loci are reported in Table S2).

2.6 | Analysis of population genomics data

The genetic variability within population samples was assessed on polymorphic loci by computing the observed heterozygosity (H_o) and unbiased expected heterozygosity (H_e) with GENETIX 4.05.2 (Belkhir, Borsa, Chikhi, Raufaste, & Bonhomme, 2000). Allelic richness (A_R) was calculated using HP-RARE (Kalinowski, 2005) based on the smallest sample size (N = 16) across all population samples.

The analyses of population differentiation and population genetic structuring were performed using different programs. ARLEQUIN 3.5 (Excoffier & Lischer, 2010) was used to calculate pairwise genetic distances between populations (F_{ST}), as well as nonhierarchical and hierarchical analysis of molecular variance (AMOVA; Excoffier, Smouse, & Quattro, 1992); in the latter case, all the possible partitions of populations in two and three groups were tested. Considering the presence of missing data, due to our filtering strategy, all the analyses were performed using the locus-by-locus option in ARLEQUIN. Significance levels for multiple comparisons were adjusted according to Benjamini and Hochberg (1995) correction for multiple tests. The package ADEGENET for R 3.2.3. (Jombart, 2008) was used for discriminant analysis of principal components (DAPC), a multivariate method to represent clusters of genetically related individuals providing a useful visual assessment of between-population differentiation.

Given the genetic homogeneity found in the Adriatic–Ionian basins (see Section 3), we carried out forward-time simulations with SimuPOP (Peng & Kimmal, 2005) using different mutation rates and population sizes, to determine the number of generations and the amount of gene flow needed to achieve the F_{ST} observed in the empirical dataset. We modeled eight ideal populations using the overall SNP frequencies of the Adriatic–Ionian samples. A universal standard mutation rate (2.5×10^{-8} mutation per generation; Pontes et al. 2015) and a faster mutation rate (1×10^{-6}) were tested. A wide range of population sizes was used (100–10,000), including the smallest effective population size (N_e) value estimated from *P. lividus* microsatellites temporal variation (100 individuals; Calderón, Palacín, & Turon, 2009), new estimates obtained from our SNPs with the linkage disequilibrium method by NeEstimator (Do et al., 2014; Table S4) and arbitrarily high values. Two alternative scenarios of isolation and migration were tested. First, the simulations were run under a pure drift model (0 migrants per generation), and the number of generations of divergence needed to reach and exceed the observed F_{ST} was recorded. Second, the simulations were performed under a scenario of migration; in this case, the number of generations and the number of migrants per generation needed to reach an equilibrium F_{ST} value close to the observed one were recorded. The complete set of parameters for the different scenarios is reported in Table S5.

2.7 | Outlier detection

Population genetics analyses typically assume that the markers employed are selectively neutral. To detect loci under directional selection, we used the F_{ST}-outlier method implemented in LOSITAN (Antao, Lopes, Lopes, Beja-Pereira, & Luikart, 2008) and the Bayesian approach of BAYESCAN 2.01 (Foll & Gaggiotti, 2008). LOSITAN was run with the following settings: 1 million simulations under neutral mean F_{ST}, confidence interval of 0.95%, a false discovery rate (FDR) of 0.01, and the infinite allele model; for each run, three replicates were performed. BAYESCAN was run with burn in = 50,000, thinning interval = 30, sample size = 5,000, number of pilot runs = 50, length of pilot runs = 5,000, and the same false discovery rate (FDR) threshold set in LOSITAN (0.01). In both cases, all the polymorphic loci were used and the comparisons were performed between Adriatic–Ionian population samples and the two additional population samples from the Western Mediterranean Sea, as well as among Adriatic–Ionian population

samples. The allele sequences of all loci identified as putative outliers were searched against a *P. lividus* transcriptome database of 188,000 contigs (NCBI database, accession number GCZS00000000.1; Gildor, Malik, Sher, Avraham, & Ben-Tabou de-Leon, 2016) using a local blastn with an e-value cutoff of 3×10^{-6}. The matching contigs of *P. lividus* transcriptome were compared to the NCBI nucleotide collection, and the top blastn hits were selected.

2.8 | Lagrangian simulations of larval dispersal in the Adriatic and Ionian basins

Potential connectivity of *P. lividus* among the eight Adriatic and Ionian locations at which genetic samples were collected was assessed by Lagrangian simulations, using a biophysical model developed to investigate larval dispersal in the Adriatic–Ionian basins (Melià et al., 2016; Schiavina, Marino, Zane, & Melià, 2014). As a thorough investigation of connectivity across the whole Adriatic–Ionian basin would require basin-wide habitat mapping, which is not available at present for *P. lividus*, we adopted a conservative approach and assessed only direct connections among the limited number of locations used for genetics. The existence of additional unmodeled *P. lividus* populations, due to the availability of suitable rocky shore habitats across the surveyed area, is expected to further increase the potential for dispersal among the modeled locations through stepwise processes, making our estimates of connectivity robust (and estimates of isolation questionable).

The oceanographic engine of the biophysical model is based on the ocean circulation dataset produced by the AREG model (Guarnieri, Oddo, Pastore, Pinardi, & Ravaioli, 2010; Oddo, Pinardi, Zavatarelli, & Coluccelli, 2006) and provided by the Adriatic Forecasting System (http://oceanlab.cmcc.it/afs). Daily average fields of current velocity and temperature cover the whole Adriatic basin down to the 39°N parallel in the Ionian Sea over a regular horizontal grid with a resolution of 1/45° (about 2.2 km) and 31 vertical sigma layers. Bathymetry has a horizontal resolution of 1/60°, and the coastline is set in correspondence with the 10-m isobath.

The biological component of the model was based on the available knowledge about key life-history traits affecting larval dispersal. Considering an age at capture for sampled individuals that varies between 0 (i.e., newly recruited) and 9 years (i.e., maximum age; Crapp & Willis, 1975), our simulations spanned the period when they were likely born, that is 2004–2013. Spawning was assumed to occur daily between April and July (Sellem & Guillou, 2007), except in days when water temperature exceeded 18°C (Spirlet, Grosjean, & Jangoux, 1998). One thousand Lagrangian particles (representing individual larvae) were released each spawning day from each sampling location. Released particles were uniformly distributed along a depth from 0.5 to 10 m (Boudouresque & Verlaque, 2001) at random positions distributed according to a 2-D Gaussian spatial distribution centered at each location with a standard deviation of 1 km. Pelagic larval duration was assigned at birth to each individual and drawn from a Gaussian distribution with mean 30 (±5) days (Fenaux et al., 1985; Pedrotti, 1993). Trajectories were stepped forward at fixed depth via an explicit fourth-order Runge–Kutta integration method with a very fine temporal step (6 min) and followed until the end of the pelagic phase or larval death. Larval mortality was considered to be triggered by water temperature: above 18°C, we introduced a mortality rate $\mu_L = 3\ d^{-1}$, corresponding to a survival probability $\sigma_L = 5\%$ per day (Privitera et al., 2011). Connectivity effectiveness (*sensu* Melià et al., 2016) was calculated as the ratio (averaged over the simulation period) between the number of particles successfully moving from one location to another and the number of particles released from the location of origin. Larvae were considered to successfully colonize the destination location if they survived their pelagic phase and their final position fell within a circular buffer with a 5-km radius from the destination location. Connectivity persistence (measuring the continuity of the flux throughout the years, *sensu* Melià et al., 2016) was defined as the stabilization coefficient (that is the reciprocal of the coefficient of variation) of the flux calculated over the simulation period. Time series of release, survival, and success rates were tested for possible trends over the 10 years of simulation via a modified Mann–Kendall test (Hamed & Rao, 1998).

2.9 | Comparison with published cytb data

As mentioned in Introduction, Penant et al. (2013) provided evidence for significant differentiation on a small geographic scale. For Adriatic and Ionian samples, differentiation was found by reanalyzing published mitochondrial cytb sequences (Maltagliati et al., 2010). Differences were stronger using frequency-based F_{ST} than using φ_{ST}, possibly reflecting strong drift effects at this marker that would produce significant F_{ST} in few generations (Penant et al., 2013).

Considering that no differences were found for Adriatic and Ionian samples in our study using SNPs (see Section 3), and the opposite was found in Penant et al. (2013) we evaluated the statistical power to detect predefined levels of genetic differentiation in the two studies. To this end, we used the forward-time simulation method of Ryman and Palm (2006), implemented in the software POWSIM 4.1, which allows simulation of sampling from populations at various levels of expected divergence using a classical Wright–Fisher model of drift without migration or mutation. We performed two sets of simulations with POWSIM, one using the cytb haplotype frequencies and sample sizes of the seven Adriatic–Ionian populations ($N = 10$ each) reported in Penant et al. (2013), and the other based on the SNPs frequencies and sample sizes of our eight Adriatic–Ionian population samples. We used a range of population sizes to estimate the number of generations needed, in the case of pure genetic drift, to produce the global F_{ST} value observed with the cytb sequences and the power to detect differences with this marker. We compared these results with those obtained with our SNPs panel using the same number of generations but a fourfold greater effective population size compared with that used for mtDNA, as expected in the case of a balanced sex ratio for diploid markers (Hedrick, 2011). For each run, 200 replicates were performed. In the case of SNPs, the executable Powsim_b was used to accommodate more than 1,000 loci.

3 | RESULTS

3.1 | DNA extraction and species identification

The extraction protocol allowed successful extraction of intact, RNA-free genomic DNA from all 275 collected samples. A total of 119 partial cytb coding sequences of about 1,000 bp were obtained by Sanger sequencing; the blastn search against the NCBI nucleotide database confirmed the identification as *Paracentrotus lividus* of all tested specimens.

3.2 | Sequencing results, filtering, and selection of loci for genetic analyses

A total of 1,076,343,894 demultiplexed and filtered quality reads were obtained from Illumina sequencing of the three 2b-RAD pools (N = 276; one technical replicate was included). After trimming and filtering for restriction site, a total of 992,809,025 reads of 32-bp length were retained (92%), resulting in about 3.5 million reads per individual. An average of 25,000 tags per individual was obtained, leading to an approximate coverage per locus of 140. In general, a uniform read coverage between samples for the same locus was observed, whereas variability was observed between different loci. A total of 212,260 2b-RAD tags were identified among 276 individuals, 138,745 (65%) were monomorphic and 73,515 (35%) were polymorphic. After filtering, we retained 1,122 polymorphic loci for further analysis. The technical replicates matched at 1,112 of 1,122 loci; nine of ten differences were due to missing loci between the two replicates. Of the 1,122 loci, 261 loci had only one SNP, 420 had two SNPs, and 441 had three SNPs, for a total of 2,424 SNPs. As mentioned in Section 2, when a locus had 2 or 3 SNPs, only the SNP globally showing the highest expected heterozygosity was retained. The replicated individual and three individuals showing more than 30% of missing loci were excluded a posteriori from the analysis: The remaining 272 individuals shared on average 95% of loci. Lastly, the consensus sequences of these 1,122 loci were employed in a local blastn against the mitochondrial genome of *P. lividus* (NCBI accession NC_001572; Cantatore, Roberti, Rainaldi, Gadaleta, & Saccone, 1989). As no matches were found with the mitochondrial sequences, the 1,122 loci can be considered as nuclear markers.

3.3 | SNPs validation by Sanger sequencing

The consensus sequences of 179 loci from a preliminary run were used for a local blastn search against the EST-linked sequences dataset, and a total of 28 matches were identified. The alignment of each locus and its matching EST-linked sequences were analyzed, and 10 loci with different polymorphism level were chosen for the primers design. PCR conditions were optimized for each locus, and nine loci were successfully amplified on a panel of 4–6 individuals: The PCR products of five loci were too long to be sequenced (>2,000 bp); considering that PCR primers were designed on EST-linked sequences, this result probably reflects the presence of introns within these PCR products. The remaining four loci were successfully sequenced, and the 2b-RAD genotypes were confirmed (Table S2).

3.4 | Genetic analyses

Comparable values were obtained among all the population samples for observed heterozygosity, unbiased expected heterozygosity, and allelic richness (Table S3), with one-way ANOVA showing no significant differences among samples ($p > .05$).

3.4.1 | Population structure

Considering all ten population samples from the Central Mediterranean (Adriatic and Ionian seas) and Western Mediterranean (France and Tunisia), the analysis of molecular variance showed that most of the variation (>99%) occurred within populations and only 0.63% among populations. However, the corresponding global F_{ST} was highly significant ($p < .0001$), and its 95% confidence interval (CI) did not include 0 (0.00435–0.00828), indicating the presence of genetic structuring across the study area. All sixteen pairwise F_{ST} comparisons between Adriatic–Ionian and the Western Mediterranean population samples were highly significant, with F_{ST} values ranging from 0.00663 to 0.02475 ($.0001 < p < .0023$; significant after Benjamini and Hochberg correction for multiple tests; Table 2) and 95% CIs ranged from 0.00211 to 0.03357. In the Western Mediterranean, the comparison between France and Tunisia (F_{ST} = 0.01013, p = .0021) was also highly significant after correction for multitest. Hierarchical AMOVA showed that the genetic variation can be partitioned into three geographic groups, the first including the eight Adriatic–Ionian population samples, the second including only the population sample from France, and the third one including the sample from Tunisia (F_{CT} = 0.01334; $p < .0001$); also in this case, the 95% CI did not include 0 (0.0089–0.01823). This subdivision into three different geographic pools maximized the amount of variation among groups with respect to any alternative configuration including either two or three groups. Notably, the subdivision into two groups as proposed in the literature (Central vs. Western Mediterranean, Maltagliati et al., 2010) was still significant, although with a smaller F_{CT} (F_{CT} = 0.01101; $p < .0001$). This genetic pattern was confirmed by DAPC (Figure 3), which shows that samples pooled into three groups: one comprising all the samples from the Central Mediterranean, one associated with the population sample from France, and one including the population sample from Tunisia.

At the basin scale, by contrast, pairwise F_{ST} values among the eight Adriatic–Ionian population samples were all negligible (F_{ST} ranging from −0.00075 to 0.00559) and statistically not significant after multiple test correction (Table 2, gray cells). The AMOVA conducted on the Adriatic–Ionian population samples confirmed the pattern of genetic homogeneity at the basin scale (F_{ST} = 0.00217, p = .4171).

SimuPOP forward-time simulations revealed that, in the case of complete isolation under pure drift, the observed F_{ST} of Adriatic–Ionian population samples can be reached and exceeded in a few generations (Table S5). In particular, a maximum of 45 generations

TABLE 2 Pairwise genetic distances (F_{ST}) between Adriatic–Ionian samples and samples from France and Tunisia based on 1,122 polymorphic loci. Benjamini & Hochberg correction for multiple tests was applied. F_{ST} indices and *p*-values are reported below and above the diagonal, respectively; significant indices in bold. Comparisons between Adriatic–Ionian populations in gray. See Table 1 for location acronyms

	OTH	KAP	BOK	KOR	TRE	TOG	OTR	POC	FRN	TUN
OTH		.2658	.3513	.7535	.2974	.0343	.0562	.4461	**<.0001**	**<.0001**
KAP	.00244		.0707	.9170	.5003	.8506	.4594	.3562	**.0023**	**<.0001**
BOK	.00283	.00508		.1430	.3254	.5018	.3289	.0690	**<.0001**	**<.0001**
KOR	.00053	−.00064	.00507		.8482	.5956	.8599	.6084	**<.0001**	**<.0001**
TRE	.00257	.00166	.00329	.00015		.2869	.4514	.9383	**<.0001**	**<.0001**
TOG	.00495	.00003	.00245	.00170	.00288		.8191	.8147	**<.0001**	**<.0001**
OTR	.00472	.00203	.00353	.00038	.00233	.00061		.6421	**<.0001**	**<.0001**
POC	.00180	.00227	.00559	.00151	−.00075	.00050	.00147		**.0001**	**<.0001**
FRN	**.01264**	**.00663**	**.01633**	**.01122**	**.01121**	**.00984**	**.01488**	**.01093**		**.0020**
TUN	**.02475**	**.01999**	**.01886**	**.02029**	**.01612**	**.02350**	**.01882**	**.02233**	**.01013**	

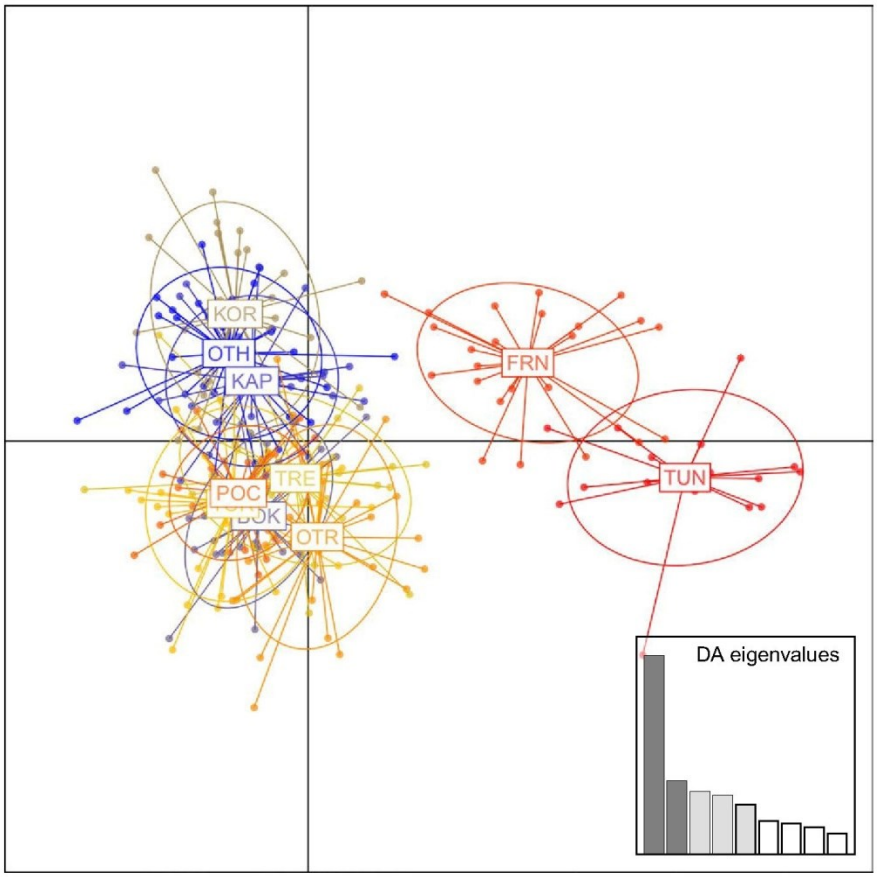


FIGURE 3 Discriminant Analysis of Principal Components (DAPC) performed by package ADEGENET. See Table 1 for location acronyms

was necessary to reach and exceed the Adriatic–Ionian F_{ST} (0.00217). This maximum value was obtained using a population size of 10,000, whereas considering a value of 100 (N_e from microsatellite temporal variation; Calderón, Palacín, et al., 2009) and 1,000 (the biggest N_e calculated from our data; Table S4), the observed F_{ST} was obtained in as few as 1–5 generations. On the other hand, in a scenario of migration, 5–15 migrants were sufficient to achieve and consistently maintain the observed F_{ST}. In this case, the equilibrium was reached in as few as 2–15 generations when the population size was set to 100–1,000 individuals, or in a maximum of 120 generations for extreme, perhaps unrealistic, population sizes (Table S5). All these results were not affected by the use of different mutation rates.

3.4.2 | Outlier detection

To investigate whether the differentiation signal was generated by one or few loci putatively under positive selection, the dataset of 1,122 polymorphic loci was explored both with LOSITAN

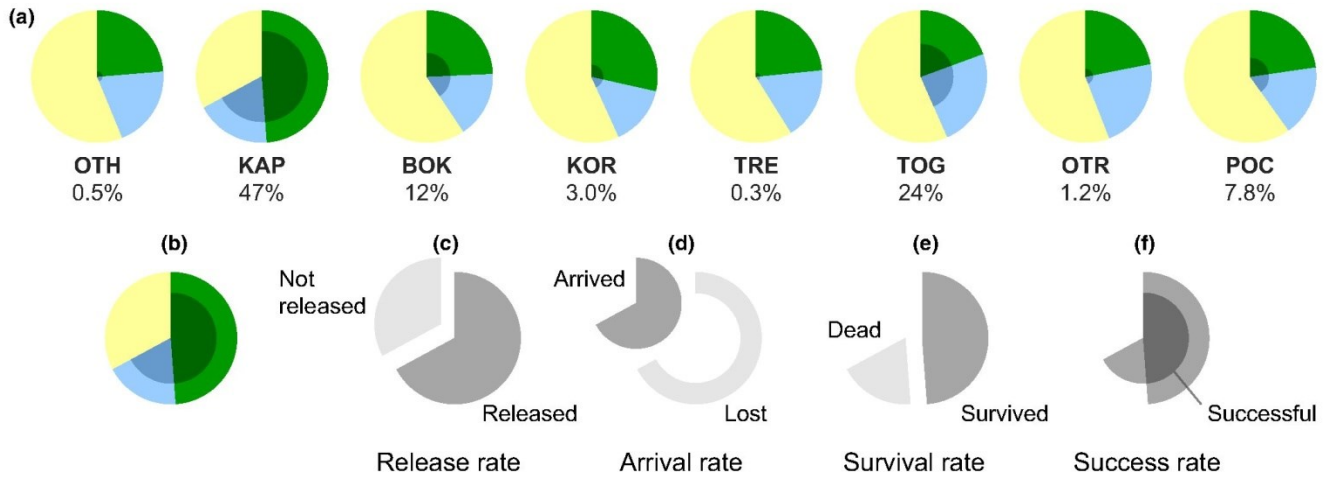


FIGURE 4 Geographic variation of connectivity among sampling locations as obtained from the biophysical model. The pie charts in the first row (a) detail the factors determining the success rate associated with each sampling location (See Table 1 for location acronyms), expressed as the percentage of successful larvae over the actual release, averaged over the whole simulation period. Pies in the second row (b–f) serve as a legend, using the KAP pie (b) as an example. The actual fraction of larvae released from a sampling location (c, release rate) depends on the presence of favorable thermal conditions for spawning (see text). Only a fraction of the larvae that are actually released reach other locations (d, arrival rate) or survive the dispersal phase (e, survival rate). The fraction of successful larvae dispersing from the location of release to any other location (f, success rate) is eventually obtained as the intersection between larvae arrived (dark gray slice in d) and larvae survived (dark gray slice in e)

and BAYESCAN. After FDR correction and considering the results obtained in all three replicates, LOSITAN identified a total of 17 loci as putative outliers under positive selection (high F_{ST}/H_E ratio) when comparing Central and Western Mediterranean population samples (Table S6). Specifically, five outliers were identified between Adriatic–Ionian population samples and France, and 12 between Adriatic–Ionian samples and Tunisia. No outliers were detected when comparing Adriatic–Ionian population samples. Conversely, BAYESCAN did not identify any locus as a putative outlier under positive selection. Interestingly, nine of the 17 putative outliers identified by LOSITAN had a match with the transcriptome of *P. lividus*, showing 93%–100% of query identity (30–32 bp). Due to some redundancy in *P. lividus* transcriptome, in six of the nine cases (asterisk in Table S6), the allelic sequences of putative outliers displayed a match with more than one contig, but these were identical and even so all alleles of each outlier pointed to the same contig hits. Six of these nine transcripts displayed a top blastn hit with *Strongylocentrotus purpuratus* proteins in the NCBI nucleotide database, and notably one of these transcripts revealed 99% query coverage and 76% identity with the codifying sequence of the hyalin gene (Table S6).

The pattern of genetic structure obtained with all 1,122 loci was confirmed also when excluding the 17 putatively selected outliers from the dataset (Table S7). For instance, compared to the dataset comprising all 1,122 loci, a slight reduction in the variation among the 10 populations was found with the 1,105 putative neutral loci (global F_{ST} = 0.00551, p < .0001). Similarly, the hierarchical AMOVA on the three groups identified above (Adriatic–Ionian samples, France, and Tunisia) confirmed the genetic structure obtained with the full dataset (F_{CT} = 0.01067, p < .0001). The pairwise F_{ST} between Central and Western Mediterranean samples and between France and Tunisia remained small but highly significant after correction for multitest, and comparisons among Adriatic–Ionian samples were all nonsignificant.

3.5 | Lagrangian simulations of larval dispersal in the Adriatic and Ionian basins

The strength of each sampling location to serve as a direct source of propagules for other locations is summarized through pie diagrams in Figure 4. The proportion of actual released larvae (Figure 4) corresponds to the fraction of the reproductive period in which water temperature in the source location was favorable to spawning (see Section 2). This fraction ranged between 40% (POC) and 67% (KAP), with an average across locations equal to 46%. Survival rate (i.e., the fraction of larvae that were not killed by mortality events triggered by water temperatures >18°C) reflected thermal variation experienced by larvae during their dispersal and ranged between 20% (TOG) and 49% (KAP) of the actual release from each location, with an average across locations of 27%. Success rate (i.e., the fraction of released larvae which survived until reaching a destination location at the end of their PLD) was on average 7% of the actual release and ranged, with a conspicuous variation spanning orders of magnitude from location to location, between 0.3% (TRE) and 47% (KAP).

Estimates of connectivity effectiveness (Table 3) and persistence (Table 4) for successful particles, as derived from Lagrangian simulations, are reported in matrices whose rows/columns refer to the eight sampling locations of the Adriatic–Ionian seas (see again Figure 2) ordered in a counterclockwise direction from OTH to OTR. Retention (i.e., self-connectivity, represented by the diagonal elements of the matrix) was lowest (<1%) in three locations (OTH, TRE, OTR), intermediate (between 1% and 20%) in two (BOK, POC), while it was highest

	OTH	KAP	BOK	KOR	TRE	TOG	OTR	POC
OTH	0.224	0.269	0.002	–	–	–	0.005	0.015
KAP	0.029	47.077	0.036	–	–	–	0.007	0.001
BOK	–	0.001	12.446	0.003	–	–	–	–
KOR	–	0.000	–	2.997	0.006	–	–	–
TRE	–	0.000	–	0.019	0.002	0.197	0.054	0.043
TOG	–	0.002	0.003	–	–	23.869	0.046	0.030
OTR	0.001	0.063	0.001	–	–	0.003	0.660	0.474
POC	–	–	–	–	–	–	–	7.759

TABLE 3 Connectivity effectiveness for *Paracentrotus lividus*. Connectivity effectiveness (estimated via Lagrangian simulation) is measured as the proportion (averaged over the simulation period) of larvae successfully moving from the locations of origin (in rows) to the destination locations (in columns) with respect to the potential releases. Positive values are reported as percents. Shaded cells indicate retention (i.e., self-connectivity). See Table 1 for location acronyms

	OTH	KAP	BOK	KOR	TRE	TOG	OTR	POC
OTH	1.326	0.651	0.444	–	–	–	0.441	0.649
KAP	0.671	2.293	0.357	–	–	–	0.667	0.433
BOK	–	0.316	3.962	0.463	–	–	–	–
KOR	–	–	–	1.284	0.316	–	–	–
TRE	–	–	–	0.809	0.316	1.236	1.177	0.384
TOG	–	0.454	0.316	–	–	1.914	0.962	0.834
OTR	0.316	0.465	0.316	–	–	0.394	1.365	0.824
POC	–	–	–	–	–	–	–	2.039

TABLE 4 Connectivity persistence for *Paracentrotus lividus*. Connectivity persistence is defined as the stabilization coefficient (i.e., the reciprocal of the coefficient of variation) of the average annual flux of larvae successfully moving from the locations of origin (in rows) to the destination locations (in columns), calculated over the simulation period. Shaded cells indicate retention. See Table 1 for location acronyms

(>20%) in TOG and KAP. Self-connections had also higher levels of persistence compared to cross-connections, with the exception of TRE. All sampling locations were connected to the nearest location in the counterclockwise direction, as revealed by the first elements above the diagonal in the matrix of connectivity effectiveness (Table 3). Those elements were systematically larger than the corresponding ones below the diagonal (which represent connections in the opposite direction); the only exception was the connection between TRE and KOR, in which the exchange of particles from west to east was stronger (in terms of both effectiveness and persistence) than that from east to west. It is apparent that only a few off-diagonal elements of the matrices in Tables 3 and 4 are not null. In terms of larval connectivity for *P. lividus* in the focal area, this means that only in some cases do the propagule fluxes directly connect sampling locations that are not nearest neighbors. Particles originating from Greek coasts (OTH) were able to reach Montenegro (BOK), and particles from both OTH and KAP were successful in crossing the Adriatic Sea and reaching some Italian sampling locations (OTR and POC). All Italian locations were connected with each other from north to south (TRE to POC), in most cases through persistent (>0.75) connections. TRE acted as an important source of particles for all Italian locations, while POC acted as an effective receptor of particles coming from most locations on both sides of the basin (with the exception of the two northernmost locations on the eastern side, BOK and KOR). Some particles were occasionally able to cross the Adriatic Sea from west to east, starting from TOG and OTR.

Changes in connectivity throughout the reproductive season following the effect of changing water temperature are shown in Figure 5. Release rate (averaged across all the release locations) decreased from April to July (approximately from 99% to 6%, Figure 5a). It was quite similar in all locations but KAP, where lower water temperatures determined release rates as high as 40% also in June and July. Survival rate (which is affected by the temperature experienced by particles along their trajectories) followed a nonlinear pattern over the reproductive season, with a maximum in April (83%) and a minimum in May (26%). Success rate, as a result of the interaction between seasonal variation in water temperature and current patterns, varied between a minimum of about 8% in May and a maximum in June and July (44%) due to high rates of particle arrival in summer months. The effect of temperature on the connectivity effectiveness matrix is shown in Figure 5b. Connectivity patterns were richer in April and May, while in June and July they were reduced to a few self-connections on the eastern side of the basin. The analysis of interannual variation of release, survival, and success rates showed no significant trend in release and success rate ($p > .2$), while survival showed a significant ($p = .01$) negative trend ($y = -0.015t + 31.05$), suggesting that changes in water temperature may have determined a reduction in larval survival over the last decade (Fig. S1).

3.6 | Comparison with published cytb data

The available cytb sequences for Adriatic and Ionian samples (seven population samples from Maltagliati et al., 2010; Penant et al., 2013) provided a F_{ST} value of 0.02222, 10 times higher than the F_{ST} calculated for the SNPs dataset ($F_{ST} = 0.00217$). POWSIM analysis (Table 5) showed that the cytb F_{ST} value can be obtained, in the case

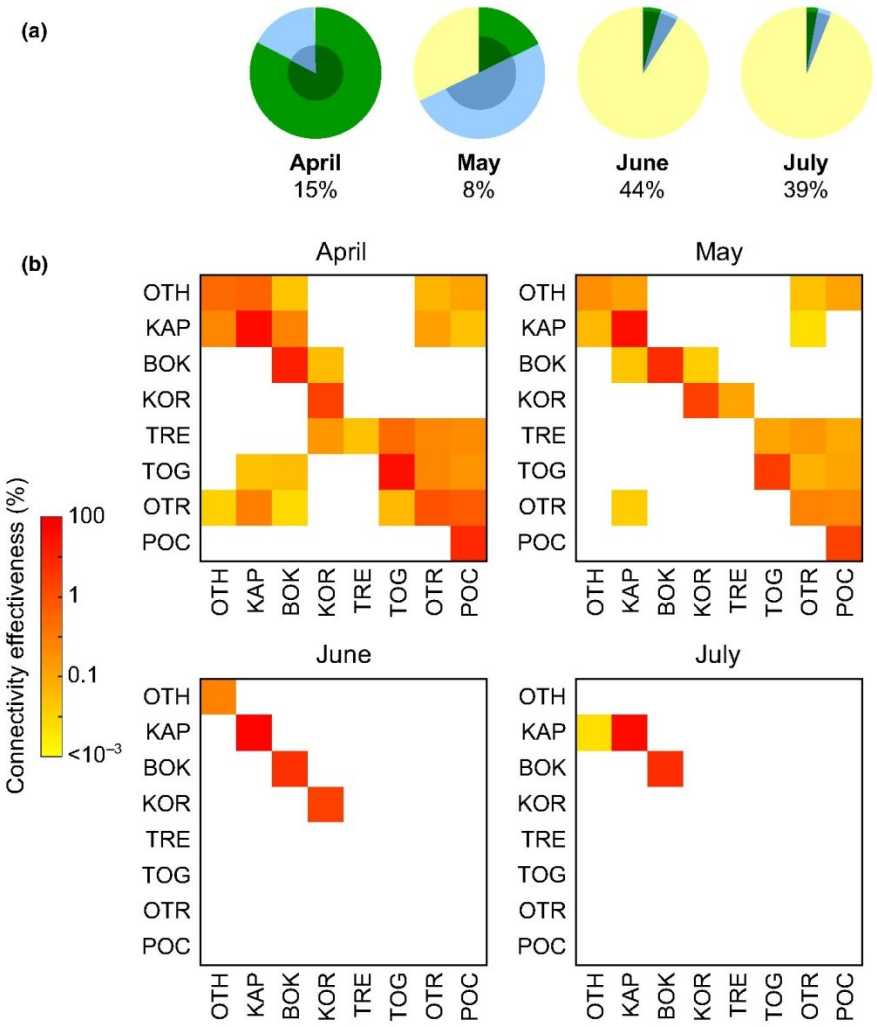


FIGURE 5 Monthly variation in potential connectivity as obtained from the biophysical model. (a) Success rate over the actual release (averaged over years and sampling locations; see Fig. 4 for color codes). (b) Connectivity effectiveness matrices (averaged over years; see Table 1 for location acronyms)

of complete isolation, in as few as 2–5 generations assuming a population size of 100 females, or in a maximum of 100–120 generations when considering the highest female population size of 2,500 used in our forward-time simulations. For all the combinations of generations and female population sizes, the power to detect differences with mtDNA was rather low (range 7.5%–39%), reflecting the small size of the population samples in the original study (N = 10 each). By contrast, the power of our panel of SNPs estimated by using the same number of generations and a fourfold greater diploid effective population size (as expected in the case of an equal sex ratio), was very high, approaching 100% in the majority of the cases. This result clearly indicates that the detection of differences between Adriatic–Ionian samples with cytb sequences but not with our panel of SNPs is very unlikely under the simulated conditions.

4 | DISCUSSION

Overall, the 2b-RAD population genomics analysis of the eight Adriatic–Ionian and two French-Tunisian population samples of *Paracentrotus lividus* examined in this study provide clear support for genetic differentiation between the Central and Western Mediterranean Sea and for differentiation between samples collected along the French and Tunisian coasts. In the Central Mediterranean, genetic homogeneity was detected within and between Adriatic and Ionian basins, despite the use of more than 1,000 polymorphic loci. Lagrangian simulations in the Central Mediterranean (Adriatic and Ionian seas) predicted a potential larval exchange among locations, altogether supporting the connectivity of *P. lividus* in this area from a seascape genetics perspective.

4.1 | Genetic differentiation at the large scale

Our results strongly support evidence of genetic partitioning among population samples collected throughout the entire study area, with a clear genetic break separating Central (Adriatic–Ionian) and Western (France–Tunisia) Mediterranean samples. This result is consistent with previous studies (Maltagliati et al., 2010; Penant et al., 2013), based mainly on mtDNA sequencing, where significant genetic divergence between the Adriatic Sea and the rest of the Mediterranean was detected. Notably, the pairwise F_{ST} values between Western and Central Mediterranean samples, although highly significant, were

TABLE 5 Statistical power of SNPs in comparison with published cytb data. POWSIM simulations were performed on mitochondrial data using the cytb haplotype frequencies of 70 sequences from 7 Adriatic–Ionian populations (Santa Caterina di Nardò, Brindisi, Manfredonia, Lesina, Ancona, Mljet, Miramare) obtained originally in Maltagliati et al. (2010) and on nuclear SNPs using our eight Adriatic–Ionian samples. For cytb simulations, several values of female haploid effective population size (N_f) were used in combination with a different number of generations of pure drift, to simulate a F_{ST} value of 0.0222, similar to that observed in Maltagliati et al. (2010) samples. Reported are: the N_f value used, the number of generations of drift (t), the average F_{ST} value obtained from 200 replicates and the power to detect differentiation, calculated as the proportion of significant tests at the end of the simulations using the Chi-square test. For SNPs simulations, the same number of generations was tested using a diploid effective population size (N_e) four times larger than the corresponding N_f. The N_e values, number of generations of drift (t), average nuclear F_{ST} value obtained from 200 replicates and the power to detect differentiation, calculated as the proportion of significant tests at the end of the simulations using the Chi-square test are reported

Cytb				SNPs			
N_f	t	F_{ST}	Power	N_e	t	F_{ST}	Power
100	2	0.0100	0.075	400	2	0.0025	1
	5	0.0249	0.385		5	0.0062	1
500	20	0.0198	0.245	2,000	20	0.0050	1
	25	0.0245	0.390		25	0.0062	1
1,000	40	0.0197	0.260	4,000	40	0.0050	1
	50	0.0250	0.380		50	0.0062	1
2,500	100	0.0198	0.195	10,000	100	0.0050	1
	120	0.0236	0.320		120	0.0069	1

small, indicating that shallow differentiation exists also at this very wide geographic scale (>2,000 km).

In addition, the 2b-RAD approach was able to detect a further, although weak, north-to-south differentiation in the Western Mediterranean, despite the relatively small sampling effort. This differentiation emerged both from pairwise comparisons and a hierarchical analysis of molecular variance supporting a subdivision of the samples into three geographic groups (Adriatic–Ionian basins, France, Tunisia) against any alternative partitioning including only two groups.

One of the most important advantages provided by genome-wide approaches is the opportunity to disentangle the signal of differentiation obtained with neutral loci and with loci putatively under selection, potentially providing complementary information that could be used in defining conservation units (Funk, McKay, Hohenlohe, & Allendorf, 2012). The genomic scan analysis of sea urchin samples identified 17 putative outliers under directional selection. However, these loci were identified only by LOSITAN, which is very sensitive to deviations from a basic island model (Lotterhos & Whitlock, 2014; Whitlock & Lotterhos, 2015), whereas BAYESCAN did not detect outliers, raising the possibility that these outlier loci could simply be the product of genetic drift and a complex demographic history. Anyway, the inclusion of these 17 putative outliers did not change the general picture of differentiation. In fact, the exclusion of these loci resulted in smaller values of differentiation compared with the overall dataset but identified the same three geographic groups. The interpretation of these putative outliers as being truly under selection is premature, especially considering our limited sampling that prevented correlation of allelic frequencies with environmental variables. At present, it is worthwhile to note that nine of the 17 putative outliers matched with the recently released transcriptome of *P. lividus*, and six of them have a correspondence with *Strongylocentrotus purpuratus* proteins in the database of NCBI. The most intriguing match is for the fibrillary glycoprotein hyalin (Table S6), a key protein involved in the formation of the fertilization membrane, which avoids polyspermy just after fertilization (McClay & Fink, 1982). This match potentially adds a new gene linked to fertilization processes, in addition to the well-known examples of gamete recognition proteins (Calderón, Turon, & Lessios, 2009; Pujolar & Pogson, 2011), for studies of sea urchins based on candidate genes.

4.2 | Connectivity at the Adriatic–Ionian sea scale

Whether or not the putative outliers are taken into account, the results of our molecular analysis highlight an incontrovertible genetic homogeneity among the eight Adriatic–Ionian population samples. Both the analysis of molecular variance and all pairwise comparisons strongly support a pattern of genetic homogeneity in this area, where all the analyzed population samples seem to belong to a single panmictic population, at least from the genetic point of view. None of the pairwise F_{ST} comparisons was significant after correction for multiple tests, and only one pairwise comparison was below the uncorrected 0.05 probability threshold (TOG-OTH).

Remarkably, SimuPOP forward-time simulations provided clear indications that this Adriatic–Ionian genetic homogeneity is unlikely in the case of isolation, but is rather sustained by gene flow. In fact, in the case of isolation, the differentiation observed in the empirical dataset is reached and quickly exceeded in a few generations. On the other hand, forward-time simulations with migration showed that few migrants (5–15 individuals depending on the population sizes) are sufficient to quickly and stably reach the F_{ST} value obtained for the Adriatic–Ionian basins. Taking into account the population sizes used in forward-time simulations (Table S5), these small number of migrants

correspond to a migration rate ranging from 0.15% to 5%, thus to a relatively high level of connectivity.

Notably, a scenario of genetic homogeneity with gene flow is consistent with the potential for larval connectivity obtained via Lagrangian simulations, performed by taking into account species-specific life traits of spawning and dispersal. Simulations suggest that each sampling location within the Adriatic–Ionian basin is potentially connected to its closest neighbor (mainly in a counterclockwise direction, as explained above) through a relatively persistent flux of propagules. In addition to these along-shore connections between neighboring locations, a few cross-basin dispersal events (between the eastern and western coasts of the Adriatic) can take place in both directions and may sometimes reach also the furthest (Ionian) locations. In this regard, it is important to note that, according to our settings, the Lagrangian simulations should provide lower-bound (conservative) estimates of potential connectivity. This conservativeness is the result of at least three major restrictions. First, we considered only a single spawning season between April and July (according to Sellem & Guillou, 2007), even though two possible spawning periods were reported for some location in the Mediterranean (spring; Byrne, 1990; Lozano et al., 1995; late summer–autumn; Régis, 1979; Fenaux et al., 1985; Pedrotti, 1993), and there is evidence for the presence of mature individuals all year round (Guettaf, San Martin, & Francour, 2000). Second, we included a water temperature effect in our model under the assumption that spawning occurs only below 18°C (Spirlet et al., 1998) and that larval survival is strongly reduced at temperatures above this threshold (Privitera et al., 2011). Both assumptions predict reduced connectivity during summer. Third, dispersal was modeled only among locations at which genetic sampling was performed, whereas rocky shore habitats suitable for *P. lividus* are widely available across the surveyed area. It is therefore very likely that stepwise, along-shore connections between unsampled locations can greatly increase the potential for dispersal among the modeled locations. For all these reasons, we believe that our general picture of potential dispersal, suggesting a general mixing within the Adriatic–Ionian seas, provides support for the hypothesis that the genetic homogeneity observed in the area is due to exchange between locations.

4.3 | Discrepancies with published cytb results

The absence of genetic structure within the Adriatic–Ionian basins obtained with SNPs is in sharp contrast with the significant differentiation detected in the same area using mtDNA sequences (Penant et al., 2013). Interestingly, in their study the authors obtained strong and significant differentiation for the Adriatic–Ionian population samples by reanalyzing published cytb data (Maltagliati et al., 2010) using pairwise F_{ST} values, and thus considering allelic frequency only. More generally, Penant et al. (2013) highlighted a much higher level of differentiation by using F_{ST} rather than taking into account genetic distances (Φ_{ST}), with significant F_{ST} values among "nearly all population pairs within the Mediterranean and Adriatic basins, some of which were geographically close (about 40–60 km between our populations from the French Mediterranean coast)" and Φ_{ST} values capturing only differences at major transitions. This pattern was explained by a lack of informativeness of mtDNA mutations at the local scale, contrasted by the possibility to obtain significant F_{ST} values after relatively few generations of genetic drift (Penant et al., 2013). However, our POWSIM forward-time simulations performed on cytb and SNPs data strongly point toward a more complex picture in the Adriatic and Ionian area. In fact, under a wide range of simulated divergence times and effective population sizes, and assuming a 1:1 sex ratio, significant F_{ST} values should be observed with our wide panel of SNPs loci but not with cytb, which is exactly the opposite of what we observed in our study.

It is important to note that these simulations rule out the possibility that an undetected unbalanced effective sex ratio can be at the basis of the different patterns observed with the two set of markers in the Adriatic–Ionian area, because the diploid effective population size itself is limited by a small number of effective females, being bound to an upper limit of four times the female population size in the case of a standard mating system (Wright, 1931).

Our forward-time simulations showed the high power of our SNPs panel to detect differences in a few generations of isolation. For this reason, it is unlikely that mtDNA markers can detect differences that do not emerge from nuclear markers. Thus, a higher mutation rate and degree of polymorphisms of mtDNA markers compared to SNPs, and the uniparental inheritance and ploidy level, which make mtDNA more strongly affected by genetic drift (Ballard & Whitlock, 2004; Weersing & Toonen, 2009), are not likely explanations for the discrepancy between the mtDNA and SNPs results.

Alternatively, the contrasting pattern could be caused by selection or, perhaps, by the strongly skewed reproductive success of females. In this latter case, the successful reproduction of some females and the subsequent recruitment of their offspring would quickly increase maternal haplotype frequency leading to inflated differentiation. In particular, mtDNA markers can be influenced by differences in sex dispersal or by mating incompatibilities driven by females. The first case is not likely when fertilization is external and adults have limited mobility, as occurs in *P. lividus*. On the other hand, intraspecific mating incompatibilities between male/female pairs have been reported in sea urchins (Kregting, Thomas, Bass, & Yund, 2014); furthermore, in *P. lividus* differences in allelic frequency of bindin, a key gene involved in sperm recognition, were detected among otherwise genetically homogeneous cohorts (Calderón & Turon, 2010; Calderón, Palacín, et al., 2009).

4.4 | Comparison with previous results in the same area and implications for MPAs

The geographic framework of this study is the selected Pilot Area of the FP7 CoCoNET project. The project aimed to investigate connectivity patterns of different species to understand the scale at which MPAs in this region can work as an effective network. Our findings support the connectivity in this area (at least for a species with a long PLD, as in the edible sea urchin). Our results are consistent with the evidence of complete homogeneity (obtained via both macrosatellites

and Lagrangian simulations) for the marbled crab *Pachygrapsus marmoratus*, a semiterrestrial crab inhabiting upper and middle tidal levels of rocky shores, with a PLD of 1 month, which was sampled approximately at the same locations (Marino et al., in review). Furthermore, our findings agree with the results obtained for the white sea bream (*Diplodus sargus sargus*), a relatively sedentary species (D'Anna, Giacalone, Pipitone, & Badalamenti, 2011; Di Franco et al., 2011) with a 1-month PLD, for which an overall genetic homogeneity in the southwestern Adriatic Sea was observed using microsatellites (Di Franco et al., 2012; Pujolar et al., 2013). Based on this evidence, rocky shore populations of species with a high dispersal potential can be considered as effectively connected from the genetic point of view within the Adriatic–Ionian geographic region.

This suggests that a trans-border approach (involving Italy, Croatia, Montenegro, Albania, and Greece) to the planning and management of MPA networks might be necessary. However, from a management perspective, it is essential to consider that the lack of a detectable genetic differentiation does not necessarily imply that *P. lividus* populations should be managed as a single unit, because the amount of migration needed to maintain genetically homogeneous populations is far smaller than the amount required to make them indistinguishable from a demographic viewpoint (Bortolotto, Bucklin, Mezzavilla, Zane, & Patarnello, 2011). Considering the threshold suggested by Waples and Gaggiotti (Waples & Gaggiotti, 2006), up to 10% of individuals can be exchanged at each generation among populations maintaining their demographic independence. In our simulations with migration (Table S5), the level of migration between populations (0.15%–5% per generation) was lower than 10% of individuals with any value of N_e. Assuming that these values of N_e are sensible, the level of exchange between our samples is not high enough to justify their management as a single population. In contrast, taking into account the potential for self-seeding suggested by the Lagrangian simulations (Tables 3 and 4), and the potential for larval export suggested by the genetic analyses, a local (yet coordinated) management of marine reserves may provide an effective conservation strategy that could also benefit the surrounding unprotected areas.

On the other hand, the comparison with other results recently produced in the context of the CoCoNET project (all referring to the same pilot study area) indicates that genetic homogeneity is not so common. For instance, while multispecies biophysical modeling (Melià et al., 2016) suggested a quite high potential connectivity for the black scorpionfish (*Scorpaena porcus*), a rocky shore fish characterized by a PLD of about 1 month, microsatellites genetic analyses pointed out a clear subdivision for the same species between the eastern and western part of the Adriatic and Ionian basins (Boissin et al., 2016). In addition, when considering a key ecological species, the habitat former *Posidonia oceanica*, whose propagules float passively for about 28 days, microsatellites showed a relatively low connectivity with a possible north-to-south partition (Jahnke et al., in review).

Although still based on a relatively small number of species, this comparison clearly warns against inferring a general pattern based only on the duration of the dispersal phase of each species. While a propagule duration of about 1 month may be sufficient to ensure genetic mixing in some species, such as *P. lividus*, it may be insufficient to produce genetic homogeneity in others. Specific dispersal traits of propagules (such as reproductive timing, depth range, behavioral features), together with differences in the spatial distribution, abundance and fecundity of their source populations, as well as recruitment success (further modulated by density-dependent effects and ecological interactions), can indeed produce very different connectivity outcomes even when PLDs are very similar.

5 | CONCLUSIONS AND FUTURE LINES OF RESEARCH

Taken together, our results provide strong support for the existence of (i) genetic differentiation in *P. lividus* at the largest geographic scale within the Mediterranean Sea, and (ii) genetic mixing in the Adriatic–Ionian seas, where a smaller geographic scale and/or peculiar oceanographic features allow for an efficient exchange among locations for a species with long-distance dispersal like *P. lividus*. However, comparison with the few published studies regarding connectivity conducted in the same area, even in species with a similar PLD, showed the existence of different patterns of genetic differentiation, ranging from complete panmixia to an east–west or north–south genetic partitioning. These findings warn against any generalization based on the presently available information and strongly indicate the need to obtain genetic connectivity data from a wider array of species, including keystone species such as habitat formers as well as rare or threatened species. At the same time, detailed biological data are needed to refine biophysical models, such as basic information about the correlation of physical and biological drivers of reproductive timing and larval mortality, which both have a strong influence on connectivity. The integration of all these aspects represents a crucial step toward providing a general framework of connectivity at the community level, and predicting expected changes in community structure in the future. It is also a crucial step toward the holistic perspective needed to achieve the Aichi Target 11 of the Convention on Biological Diversity, which aims at increasing the proportion of protected areas in the world, working as "a well-connected system" by 2020 (Brooks, Dvorak, Spindler, & Miller, 2015).

ACKNOWLEDGMENTS

The research leading to these results was supported by the European Community's 7th Framework Programme (FP7/2007–2013) under Grant Agreement No. 287844 for the project "Towards COast to COast NETworks of marine-protected areas (from the shore to the high and deep sea), coupled with sea-based wind energy potential (CoCoNet, http://www.coconet-fp7.eu/)," and by the University of Padua grant CPDA148387/14. We are thankful to the following people for their precious help for logistics and for collecting samples on the field: "Antheus srl (Lecce, Italy)"; S Bevilacqua, F Boero, S Fraschetti, and A Terlizzi (University of Salento, Italy); D Micu (National Institute of Marine Research and Development, Romania); S Reizopoulou and

MA Pancucci-Papadopoulou (HCMR, Greece); S Beqiraj (Department of Biology, Faculty of Natural Sciences, University of Tirana); and E Hajdëri (Catholic University "Our Lady of Good Counsel," Tirana). We are also thankful to L Bargelloni, M Babbucci, R Franch, and F Maroso for their invaluable help with the 2b-RAD protocol and to N Forin (Department of Biology, University of Padua) for the technical assistance. We are deeply indebted with Jilda Alicia Caccavo (Department of Biology, University of Padua) for her help with English. Useful discussions during the CoCoNet meetings with all partners were also very instrumental in the development of some of the ideas that guided our research.

DATA ACCESSIBILITY

SNPs genotypes data together with the datasets supporting the conclusions of this article are available as supporting information and additional files in DRYAD at doi:10.5061/dryad.73kc6. Short read data are deposited in SRA with accessions SRR5266202-SRR5266474.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

MP, PM, MS, and LZ conceived and designed the experiments and wrote the manuscript. MP and MS performed the experiments. MP, PM, MS, and LZ analyzed the data. PM, LZ, and TP contributed reagents/materials/analysis tools. GA, JBS, EB, RC, AC, MC, LC, GG, CK, VM, IAMM, PM, CP, TP, MP, MS, and LZ critically reviewed the manuscript; the final version of the manuscript was approved by all the authors.

REFERENCES

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Bioinformatics*, *9*, 323.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, *3*, e3376.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*, 729–744.

Barnes, D. K. A., & Crook, A. C. (2001). Implications of temporal and spatial variability in *Paracentrotus lividus* populations to the associated commercial coastal fishery. *Hydrobiologia*, *465*, 95–102.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (2000). *GENETIX 4.02, Logiciel Sous Windows TM Pour La Génétique Des Populations. Laboratoire Génome, Populations Interactions, CNRS UMR 5000*.

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*, 289–300.

Boissin, E., Micu, D., Janczyszyn-Le Goff, M., Neglia, V., Bat, L., Todorova, V., … Planes, S. (2016). Contemporary genetic structure and post-glacial demographic history of the black Scorpionfish, *Scorpaena Porcus*, in the Mediterranean and the Black Seas. *Molecular Ecology*, *25*, 2195–2209.

Bortolotto, E., Bucklin, A., Mezzavilla, M., Zane, L., & Patarnello, T. (2011). Gone with the currents: Lack of genetic differentiation at the circum-continental scale in the Antarctic Krill *Euphausia superba*. *BMC Genetics*, *12*, 32.

Boudouresque, C. F., & Verlaque, M. (2001). Ecology of *Paracentrotus Lividus*. *Developments in Aquaculture and Fisheries Science*, *32*, 177–216.

Bowen, B. W., Bass, A. L., Muss, A., Carlin, J., & Robertson, D. R. (2006). Phylogeography of two atlantic Squirrelfishes (Family *Holocentridae*): Exploring links between pelagic larval duration and population connectivity. *Marine Biology*, *149*, 899–913.

Brooks, J. J., Dvorak, R. G., Spindler, M., & Miller, S. (2015). Relationship-scale conservation. *Wildlife Society Bulletin*, *39*, 147–158.

Byrne, M. (1990). Annual reproductive cycles of the commercial sea urchin *Paracentrotus lividus* from an exposed intertidal and a sheltered subtidal habitat on the West Coast of Ireland. *Marine Biology*, *104*, 275–289.

Calderón, I., Giribet, G., & Turon, X. (2008). Two markers and one history: Phylogeography of the edible common sea urchin *Paracentrotus lividus* in the Lusitanian Region. *Marine Biology*, *154*, 137–151.

Calderón, I., Palacín, C., & Turon, X. (2009). Microsatellite markers reveal shallow genetic differentiation between cohorts of the common sea urchin *Paracentrotus lividus* (Lamarck) in Northwest Mediterranean. *Molecular Ecology*, *18*, 3036–3049.

Calderón, I., & Turon, X. (2010). Temporal genetic variability in the Mediterranean common sea urchin *Paracentrotus lividus*. *Marine Ecology Progress Series*, *408*, 149–159.

Calderón, I., Turon, X., & Lessios, H. A. (2009). Characterization of the sperm molecule bindin in the sea urchin genus *Paracentrotus*. *Journal of Molecular Evolution*, *68*, 366–376.

Cantatore, P., Roberti, M., Rainaldi, G., Gadaleta, M. N., & Saccone, C. (1989). The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. *The Journal of Biological Chemistry*, *264*, 10965–10975.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124–3140.

Coombs, J. A., Letcher, B. H., & Nislow, K. H. (2008). Create: A software to create input files from diploid genotypic data for 52 genetic software programs. *Molecular Ecology Resources*, *8*, 578–580.

Cowen, R. K., Paris, C. B., & Srinivasan, A. (2006). Scaling of connectivity in marine populations. *Science*, *311*, 522–527.

Cowen, R. K., & Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Annual Review of Marine Science*, *1*, 443–466.

Crapp, G. B., & Willis, M. E. (1975). Age determination in the sea urchin *Paracentrotus lividus* (Lamarck), with notes on the reproductive cycle. *Journal of Experimental Marine Biology and Ecology*, *20*, 157–178.

D'Aloia, C. C., Bogdanowicz, S. M., Francis, R. K., Majoris, J. E., Harrison, R. G., & Buston, P. M. (2015). Patterns, causes, and consequences of marine larval dispersal. *Proceedings of the National Academy of Sciences*, *112*, 13940–13945.

D'Anna, G., Giacalone, V. M., Pipitone, C., & Badalamenti, F. (2011). Movement pattern of white seabream, *Diplodus sargus* (L., 1758) acoustically tracked in an artificial reef area. *Italian Journal of Zoology*, *78*, 255–263.

DeAngelis, M. M., Wang, D. G., & Hawkins, T. L. (1995). Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Research*, *23*, 4742–4743.

Di Franco, A., Coppini, G., Pujolar, J. M., De Leo, G. A., Gatto, M., Lyubartsev, V., … Guidetti, P. (2012). Assessing dispersal patterns of fish propagules from an effective mediterranean marine protected area. *PLoS ONE*, *7*, e52108.

Di Franco, A., De Benedetto, G., De Rinaldis, G., Raventos, N., Sahyoun, R., & Guidetti, P. (2011). Large scale variability in otolith microstructure and microchemistry: The case study of *Diplodus sargus sargus* in the Mediterranean Sea. *Italian Journal of Zoology*, *78*, 182–192.

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator V2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*, 209–214.

Duran, S., Palacín, C., Becerro, M. A., Turon, X., & Giribet, G. (2004). Genetic diversity and population structure of the commercially harvested sea urchin *Paracentrotus lividus* (Echinodermata, Echinoidea). *Molecular Ecology*, *13*, 3317–3328.

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, *103*, 285–298.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin Suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, *131*, 479–491.

FAO (2004). *FAO Major Fishing Areas*. Mediterranean and Black Sea (Major Fishing Area 37). CWP Data Collection. In: *FAO Fisheries and Aquaculture Department*. Rome. Updated 1 October 2004. http://www.fao.org/fishery/area/Area37/en

Fenaux, L., Cellario, C., & Etienne, M. (1985). Croissance de la Larve de l'Oursin *Paracentrotus lividus*. *Marine Biology*, *86*, 151–157.

Foll, M., & Gaggiotti, O. E. (2008). A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*, 977–993.

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*, 489–496.

Galindo, H. M., Olson, D. B., & Palumbi, S. R. (2006). Seascape genetics: A coupled oceanographic-genetic model predicts population structure of Caribbean corals. *Current Biology*, *16*, 1622–1626.

Galindo, H. M., Pfeiffer-Herbert, A. S., McManus, M. A., Chao, Y., Chai, F., & Palumbi, S. R. (2010). Seascape genetics along a steep cline: Using genetic patterns to test predictions of marine larval dispersal. *Molecular Ecology*, *19*, 3692–3707.

Gildor, T., Malik, A., Sher, N., Avraham, L., & Ben-Tabou de-Leon, S. (2016). Quantitative developmental transcriptomes of the Mediterranean sea urchin *Paracentrotus lividus*. *Marine Genomics*, *25*, 89–94.

Gilg, M. R., & Hilbish, T. J. (2003). The geography of marine larval dispersal: Coupling genetics with fine-scale physical oceanography. *Ecology*, *84*, 2989–2998.

Guarnieri, A., Oddo, P., Pastore, M., Pinardi, N., & Ravaioli, M. (2010). The Adriatic basin Forecasting System: New model and system development. In H. Dahlin, M. J. Bell, N. C. Flemming, & S. E. Petersson (Eds.), *Coastal to Global Operational Oceanography: Achievements and Challenges* (pp. 184–190). Proceeding of the Fifth International Conference on EuroGOOS, 20–22 May 2008. Exeter, UK: EuroGOOS Publication.

Guettaf, M., San Martin, G. A., & Francour, P. (2000). Interpopulation variability of the reproductive cycle of *Paracentrotus lividus* (Echinodermata: Echinoidea) in the South-Western Mediterranean. *Journal of the Marine Biological Association of the United Kingdom*, *80*, 899–907.

Hamed, K. H., & Rao, A. R. (1998). A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology*, *204*, 182–196.

Hansen, M. M., & Hemmer-Hansen, J. (2007). Landscape genetics goes to sea. *Journal of Biology*, *6*, 6.

Hedrick, P. W. (2011). *Genetics of Populations* (p. 215). Sudbury, MA: Jones and Bartlett Publishers.

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*, 273–290.

Jahnke, M., Casagrandi, R., Melià, P., Schiavina, M., Schultz, S., Zane, L., & Procaccini, G. (in review) (resubmission to Diversity & Distribution in preparation). *Connectivity, Sexual Reproduction and Selection Pressure in the Seagrass P. oceanica and the Implications for Conservation.*

Jombart, T. (2008). Adegenet: A R Package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405.

Kalinowski, S. T. (2005). HP-RARE 1.0: A computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, *5*, 187–189.

Kregting, L. T., Thomas, F. I. M., Bass, A. L., & Yund, P. O. (2014). Relative effects of gamete compatibility and hydrodynamics on fertilization in the Green sea urchin *Strongylocentrotus droebachiensis*. *The Biological Bulletin*, *227*, 33–39.

Kroh, A., & Mooi, R. (2017). *World Echinoidea Database*. Accessed through: World Register of Marine Species at http://www.marinespecies.org/aphia.php?p=taxdetails&id=124316

Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, *23*, 2178–2192.

Lozano, J., Galera, J., Lopez, S., Turon, X., Palacin, C., & Morera, G. (1995). Biological cycles and recruitment of *Paracentrotus lividus* (Echinodermata: Echinoidea) in two contrasting habitats. *Marine Ecology Progress Series*, *122*, 179–191.

Maltagliati, F., Di Giuseppe, G., Barbieri, M., Castelli, A., & Dini, F. (2010). Phylogeography and genetic structure of the edible sea urchin *Paracentrotus lividus* (Echinodermata: Echinoidea) inferred from the mitochondrial cytochrome B gene. *Biological Journal of the Linnean Society*, *100*, 910–923.

Marino, I. A. M. M., Schiavina, M., Aglieri, G., Boscari, E., Congiu, L., Faggion, S., … Melià, P. (in review) (resubmission to *Molecular Ecology* in preparation) *Assessing Connectivity Patterns of the Marbled Crab Pachygrapsus marmoratus in the Adriatic and Ionian Seas Combining Genetic Data and Lagrangian Simulations.*

Marshall, J. J. T., & Halford, S. E. (2010). The type IIB restriction endonucleases. *Biochemical Society Transactions*, *38*, 410.

McClay, D. R., & Fink, R. D. (1982). Sea urchin hyalin: Appearance and function in development. *Developmental Biology*, *92*, 285–293.

Melià, P., Schiavina, M., Rossetto, M., Gatto, M., Fraschetti, S., & Casagrandi, R. (2016). Looking for hotspots of marine metacommunity connectivity: A methodological framework. *Scientific Reports*, *6*, 23705.

Mileikovsky, S. A. (1971). Types of larval development in marine bottom invertebrates, their distribution and ecological significance: A re-evaluation. *Marine Biology*, *10*, 193–213.

Oddo, P., Pinardi, N., Zavatarelli, M., & Coluccelli, A. (2006). The Adriatic basin forecasting system. *Acta Adriatica*, *S47*, 169–184.

Palumbi, S. R. (2003). Population genetics, demographic connectivity, and the design of reserves. *Ecological Applications*, *13*, 146–158.

Pedrotti, M. L. (1993). Spatial and temporal distribution and recruitment of echinoderm larvae in the Ligurian Sea. *Journal of the Marine Biological Association of the United Kingdom*, *73*, 513–530.

Penant, G., Didier, A., Feral, J. P., & Chenuil, A. (2013). Planktonic larvae do not ensure gene flow in the edible sea urchin *Paracentrotus lividus*. *Marine Ecology Progress Series*, *480*, 155–170.

Peng, B., & Kimmal, M. (2005). simuPOP: A forward-time population genetics simulation environment. *Bioinformatics*, *21*, 3686–3687.

Phillips, C. (2005). Using Online Databases for Developing SNP Markers of Forensic Interest. In A. Carracedo (Ed.), *Forensic DNA Typing Protocols* (pp. 83–105). Totowa, NJ: Humana Press.

Pontes, M. L., Fondevila, M., Laréu, M. V., & Medeiros, R. (2015). SNP markers as additional information to resolve complex kinship cases. *Transfusion Medicine and Hemotherapy*, *42*, 385–388. http://doi.org/10.1159/000440832

Privitera, D., Noli, M., Falugi, C., & Chiantore, M. (2011). Benthic assemblages and temperature effects on *Paracentrotus lividus* and *Arbacia lixula* larvae and settlement. *Journal of Experimental Marine Biology and Ecology*, *407*, 6–11.

Puckett, B. J., Eggleston, D. B., Kerr, P. C., & Luettich, R. A. Jr (2014). Larval dispersal and population connectivity among a network of marine reserves. *Fisheries Oceanography*, *23*, 342–361.

Pujolar, J. M., & Pogson, G. H. (2011). Positive Darwinian selection in gamete recognition proteins of *Strongylocentrotus* sea urchins. *Molecular Ecology*, *20*, 4968–4982.

Pujolar, J. M., Schiavina, M., Di Franco, A., Melià, P., Guidetti, P., Gatto, M., … Zane, L. (2013). Understanding the effectiveness of marine protected areas using genetic connectivity patterns and Lagrangian simulations. *Diversity and Distributions*, *19*, 1531–1542.

Régis, M. B. (1979). Analyse des fluctuations des indices physiologiques Chez Deux échinoides (*Paracentrotus lividus* (LMK) et *Arbacia lixula* L.) Du Golfe de Marseille. *Tethys*, *9*, 167–181.

Ruzzante, D. E., Taggart, C. T., & Cook, D. (1998). A nuclear DNA basis for shelf- and bank-scale population structure in Northwest Atlantic Cod (*Gadus morhua*): Labrador to Georges bank. *Molecular Ecology*, *7*, 1663–1680.

Ryman, N., & Palm, S. (2006). POWSIM: A computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes*, *6*, 600–602.

Sambrook, J. E., Fritsch, F., & Maniatis, T. (1990). Molecular cloning: A laboratory manual. second edition. Volumes 1, 2, and 3. Current protocols in molecular biology. Volumes 1 and 2. *Cell*, *61*, 17–18.

Schiavina, M., Marino, I. A. M., Zane, L., & Melià, P. (2014). Matching oceanography and genetics at the basin scale. Seascape connectivity of the Mediterranean Shore Crab in the Adriatic sea. *Molecular Ecology*, *23*, 5496–5507.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*, 671–675.

Selkoe, K. A., Watson, J. R., White, C., Horin, T. B., Iacchei, M., Mitarai, S., … Toonen, R. J. (2010). Taking the chaos out of genetic patchiness: Seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology*, *19*, 3708–3726.

Sellem, F., & Guillou, M. (2007). Reproductive biology of *Paracentrotus lividus* (Echinodermata: Echinoidea) in two contrasting habitats of Northern Tunisia (south-East Mediterranean). *Journal of the Marine Biological Association of the United Kingdom*, *87*, 763–767.

Spirlet, C., Grosjean, P., & Jangoux, M. (1998). Reproductive cycle of the echinoid *Paracentrotus lividus*: Analysis by means of the maturity index. *Invertebrate Reproduction & Development*, *34*, 69–81.

Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., … Slate, J. (2010). Adaptation genomics: The next generation. *Trends in Ecology & Evolution*, *25*, 705–712.

Storfer, A., Murphy, M. A., Evans, J. S., Goldberg, C. S., Robinson, S., Spear, S. F., … Waits, L. P. (2006). Putting the 'landscape' in landscape genetics. *Heredity*, *98*, 128–142.

Taylor, M. S., & Hellberg, M. E. (2003). Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. *Science*, *299*, 107–109.

Thorrold, S. R., Jones, G. P., Hellberg, M. E., Burton, R. S., Swearer, S. E., Neigel, J. E., … Warner, R. R. (2002). Quantifying larval retention and connectivity in marine populations with artificial and natural markers. *Bulletin of Marine Science*, *70*, 291–308.

Todd, C. D. (1998). Larval supply and recruitment of benthic invertebrates: Do larvae always disperse as much as we believe? *Hydrobiologia*, *375*, 1–21.

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012). 2b-RAD: A simple and flexible method for genome-wide genotyping. *Nature Methods*, *9*, 808–810.

Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, *15*, 1419–1439.

Weersing, K., & Toonen, R. J. (2009). Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series*, *393*, 1–12.

White, C., Selkoe, K. A., Watson, J., Siegel, D. A., Zacherl, D. C., & Toonen, R. J. (2010). Ocean currents help explain population genetic structure. *Proceedings of the Royal Society of London, Series B*, *277*, 1685–1694.

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of FST. *The American Naturalist*, *186*, S24–S36.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*, 97–159.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Paterno M, Schiavina M, Aglieri G, et al. Population genomics meet Lagrangian simulations: Oceanographic patterns and long larval duration ensure connectivity among *Paracentrotus lividus* populations in the Adriatic and Ionian seas. *Ecol Evol*. 2017;00:1–17. https://doi.org/10.1002/ece3.2844

Supporting Information

Table S1. 2b-RAD oligonucleotide sequences used for Illumina next-generation sequencing.

ID	Oligo	Sequences
ADPT1 ss	Anti-ILL	AGATCGGAAGAGC/3InvdT
ADAPT2 ss	5ILL-NN	CTACACGACGCTCTTCCGATCTNN
ADAPT3 ss	3ILL-NN	CAGACGTGTGCTCTTCCGATCTNN
Primer F	ILL-Mpx	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAT
Barcoded-primer R	ILL-RAD-bc	CAAGCAGAAGACGGCATACGAGAT [barcode] GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
2bRAD amp F	IC1P5	AATGATACGGCGACCACCGA
2bRAD amp R	IC2-P7	CAAGCAGAAGACGGCATACGA

Table S2. Validated loci by PCR based-approach and Sanger sequencing. Locus ID, consensus sequence, primer sequences (Forward and Reverse) and polymorphism level for each locus, sample ID and predicted genotypes for each sample are reported. SNP position is highlighted in bold (consensus sequences).

Locus ID	Consensus sequence	Primer sequences (5'-3')	Polymorphism level	Samples ID	Genotypes
1110	TGACTGTGTACAAGGATCGTGGC**G**TTGCACTG	For – CCATCTTGTCCCGAAGATGTC Rev - AGCAACCATCGGTTCCAGGAT	1 SNP	PL_CZ2	G
				PL_CZ12	G
				PL_MN26	G
				PL_CZ5	G/A
				PL_TG12	G/A
				PL_TG15	G/A
90	CTGGACCAATCAATGATGGTGGA**G**GGT**G**CATC	For - AAAACGATACCACCATGACCTC Rev - AAGGATAGTCTTTGAAACCGGTC	2 SNPs	PL_TG27	GG
				PL_OT9	GG
				PL_TR33	GG/TG
				PL_TG20	GG/GA
				PL_OT7	GG/TG
				PL_CZ5	GG/TG
3987	CCTG**C**AGCAGCAAAGCTTGTGG**A**TACATGAAG	For - GATACCCCGCCAAAGTTGAG Rev - GGTGAACTCCTGCACTTGAT	2 SNPs	PL_MN6	CA
				PL_GR26	CA
				PL_CZ30	CA/TA
				PL_PC24	CA/AA
				PL_ALB16	CA/CC
				PL_TG8	CA/CC
6408	CTCCCCCACCCAATCCCAGTGGTGACAACATA	For - GCAGTAAGGGGTCGTCCACAA Rev - GGCGTGTGAGCAGGTTGAGACC	0 SNP (monomorphic)	PL_CZ12	Consensus seq
				PL_OT23	Consensus seq
				PL_OT26	Consensus seq
				PL_MN7	Consensus seq

Table S3. Genetic variability within *Paracentrotus lividus* population samples examined in the study. The genetic variability was assessed on the total dataset (1122 polymorphic loci) and on the dataset filtered for a minor allele frequency (MAF) of 0.01 (494 polymorphic loci, asterisk). For each population sample, observed heterozygosity (H_O), unbiased expected heterozygosity (H_E) and allelic richness (A_R) are reported; allelic richness is calculated with a minimum sample size of 16 diploid individuals.

Acronym	H_O	H_E	A_R	$*H_O$	$*H_E$	$*A_R$
OTH	0.0699	0.0750	1.28	0.1509	0.1611	1.56
KAP	0.0709	0.0766	1.28	0.1523	0.1638	1.57
BOK	0.0651	0.0742	1.27	0.1423	0.1616	1.56
KOR	0.0655	0.0728	1.27	0.1424	0.1587	1.56
TRE	0.0680	0.0752	1.28	0.1455	0.1603	1.55
TOG	0.0679	0.0756	1.28	0.1446	0.1612	1.56
OTR	0.0682	0.0774	1.29	0.1455	0.1648	1.57
POC	0.0697	0.0769	1.28	0.1512	0.1669	1.58
FRN	0.0731	0.0775	1.30	0.1566	0.1654	1.59
TUN	0.0705	0.0802	1.30	0.1495	0.1698	1.59

Table S4. Estimates of Adriatic-Ionian effective population sizes based on the Linkage Disequilibrium method by NeEstimator. Given the genetic homogeneity found in the Adriatic-Ionian basins, the estimates were calculated by pooling all the Adriatic-Ionian samples. Three allele frequencies thresholds are used and, for each one, the corresponding Jackknife 95% confidence interval is reported.

	LOWEST ALLELE FREQUENCY		
	0.05	**0.02**	**0.01**
Jackknife 95% CIs	216.8 – 290.2	452.4 – 609.3	723.9 – 1009.1

Table S5. Forward-time simulations of alternative scenarios of isolation and migration obtained with SimuPOP for the 8 Adriatic-Ionian population samples. Two alternative scenarios of isolation and migration were tested: first, a pure drift model where the populations were allowed to diverge for a fixed number of generations with the number of migrants/generation (Nm) set to 0; then, a scenario of migration where, for each predefined population size, the exact number of migrants between populations per generation needed to consistently achieve the observed F_{ST} was recorded. Specifically, 8 ideal populations were simulated starting from the Adriatic-Ionian SNP frequencies; the sex ratio was assumed to be 1:1 and a random mating scheme was allowed; at each generation, parents died after the reproduction event and their offspring received the alleles by Mendelian inheritance. Different population sizes were used (100 – 10 000); this interval was chosen so as to include a small value of 100 individuals, corresponding to the smallest effective population size (Ne) estimated from *P. lividus* microsatellites temporal variation (Calderón *et al.* 2009a), values corresponding to new effective population sizes estimated with the Linkage Disequilibrium method by NeEstimator (Do *et al.* 2014; Table S4), and few arbitrarily high values. All the simulations were conducted with a universal standard mutation rate (2.5×10^{-8}; Pontes *et al.* 2015) and a faster mutation rate (1×10^{-6}). For simulations under isolation, the table reports the tested population sizes (N), the number of generations of divergence (t_d) needed to reach and exceed the observed F_{ST} in the empirical dataset (0.00217) and the final F_{ST} obtained. For simulations with migration, N is reported along with the number of generations (t_{eq}) and the number of migrants per generation (Nm) needed to stably achieve the observed F_{ST}.

ISOLATION			MIGRATION		
N	**t_d**	**Final F_{ST}**	**N**	**t_{eq}**	**Nm**
100	1	0.00520	**100**	2	5
400	2	0.00249	**400**	8	10
1000	5	0.00247	**1000**	15	12
2000	9	0.00223	**2000**	30	13
4000	18	0.00218	**4000**	60	14
10 000	45	0.00222	**10 000**	150	15

Table S6. Putative outliers under positive selection identified with LOSITAN by comparing Central (Adriatic-Ionian) and Western (France and Tunisia) population samples. Locus ID, consensus sequence, contig ID of *Paracentrotus lividus* transcriptome (NCBI database, accession number GCZS00000000.1; Gildor *et al.* 2016) and *Strongylocentrus purpuratus* proteins providing the top hit and related NCBI accession number are reported. The position of the SNP with the highest expected heterozygosity (consensus sequences) is underlined and bold. Asterisks indicate outlier loci displayed a match with more than one contig, for convenience only one contig ID is reported since they were identical.

Locus ID	Consensus sequences	*P. lividus* contig ID	*S. purpuratus* protein
342	TACCAACTTTCAAGCAAGGTGGACACA**A**ACCC	gb|GCZS01021365.1|	TBC1 domain family member 1 (LOC100891782)
16413 *	TAAATGGAAGCAAGGAATGTGGTCAAGTGA**C**G	gb|GCZS01088153.1|	cyclin-D-binding Myb-like transcription factor 1 (LOC754814)
1230 *	AGCAGTTATTCAAACCGGGTGGGAAGGA**C**GAA	gb|GCZS01098438.1|	mannosyl-oligosaccharide alpha-1,2-mannosidase isoform B (LOC581367)
3050	A**C**ATCCCCTTCAATGTACGTGGATTTCCAGTG	-	-
10028	GCACTCACTGCAAAGTGCGTGGTCATTCA**G**AA	gb|GCZS01138538.1|	myosin VII (LOC373219)
12431	TGTGAGAGTACAATCTTTGTGGT**A**GGCTGAAA	-	-
20795 *	CCATACTCATCAAATCG**G**GTGGGATCAGTCCT	gb|GCZS01130689.1|	-
21392	TCACCTGACCCAAACCTAGTGGTAG**G**TATCAT	-	-
22169 *	CTGTTCAGCTCAAGGA**C**GGTGGTATGGCTTCT	gb|GCZS01075885.1|	-
23920	GAGAATGATACAATCAGAGTGGCAAAGC**T**GTC	-	-
26862	CCAGAGACATCAAAGGCTGTGGGTTC**T**GCAAA	gb|GCZS01121683.1|	hyalin (LOC373362)
27918	ACATAATGCTCAACTTCCGTGGTCACGA**T**AAT	-	-
4345 *	GGTCTAAGGCCAATG**C**AGGTGGTACACCCATC	gb|GCZS01108283.1|	-
17408 *	TGGATCATGTCAATATTTGTGGC**A**CTTTCCTG	gb|GCZS01108741.1|	cystine/glutamate transporter (LOC584098)
19300	CAACATGGTGCAACACGAGTGGTTT**C**GATCAT	-	-
23162	CTGTGCATCTCAAGTAAGGTGGCA**C**TAATTAG	-	-
23417	GGGTATAAAACAAGCAATGTGGCCTTTGC**C**CA	-	-

Table S7. Pairwise genetic distances (F_{ST}) between Central (Adriatic-Ionian) and Western (France and Tunisia) population samples based on 1105 neutral loci. Benjamini & Hochberg correction for multiple tests was applied. F_{ST} indices and p-values are reported below and above the diagonal, respectively; significant indices and p-values in bold. Comparisons between Adriatic-Ionian populations are shaded in grey. See Table 1 in the main text for location acronyms.

	OTH	KAP	BOK	KOR	TRE	TOG	OTR	POC	FRN	TUN
OTH		0.2520	0.3405	0.7206	0.3305	0.0285	0.0555	0.4877	**<0.0001**	**<0.0001**
KAP	0.00242		0.0658	0.8763	0.5159	0.8086	0.4628	0.4101	**0.0133**	**<0.0001**
BOK	0.00278	0.00515		0.1076	0.3026	0.4549	0.2575	0.0560	**<0.0001**	**0.0002**
KOR	0.00063	-0.00031	0.00544		0.8254	0.4738	0.8065	0.5255	**0.0001**	**0.0001**
TRE	0.00239	0.00154	0.00338	0.00027		0.2924	0.5426	0.9472	**<0.0001**	**0.0023**
TOG	0.00525	0.00024	0.00269	0.00229	0.00297		0.7601	0.8072	**0.0002**	**<0.0001**
OTR	0.00477	0.00202	0.00395	0.00071	0.00192	0.00095		0.6076	**<0.0001**	**0.0001**
POC	0.00164	0.00198	0.00581	0.00194	-0.00088	0.00049	0.00161		**<0.0001**	**<0.0001**
FRN	**0.01203**	**0.00553**	**0.01655**	**0.01026**	**0.01056**	**0.00824**	**0.01391**	**0.00964**		**0.0120**
TUN	**0.02048**	**0.01543**	**0.01525**	**0.01452**	**0.01110**	**0.01818**	**0.01418**	**0.01668**	**0.00835**	

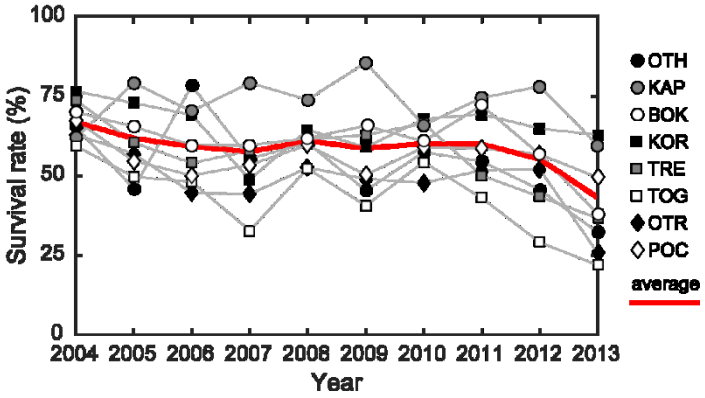


Figure S1. Survival rate trend 2004–2013. Survival rate is defined as the proportion of particles surviving their pelagic larval phase (independent of the actual success of migration from the location of origin to that of the destination). Symbols (connected by thin grey lines) indicate release locations (see Table 1 for acronyms), while the red line represents the yearly average across locations.

CHAPTER 3

Genomic population structure of the mussel *Mytilus galloprovincialis* in the Mediterranean and Black seas

Draft of the paper that will be submitted to the journal *Diversity & Distributions* as:

Paterno M, Ben Souissi J, Boscari E, Chassanite A, Congiu L, Guarnieri G, Kruschel C, Macic V, Marino IAM, Micu D, Voutsinas M, Papetti C, Zane L. **A genome-wide approach to the global phylogeography of the mussel *Mytilus galloprovincialis* in the Mediterranean and Black Seas.**

A genome-wide approach to the global phylogeography of the mussel *Mytilus galloprovincialis* in the Mediterranean and Black seas

Paterno M[1,2], Ben Souissi J[3], Boscari E[1,2], Chassanite A[4], Congiu L[1,2], Guarnieri G[2,5], Kruschel C[6], Macic V[7], Marino IAM[1,2], Micu D[8], Papetti C[1,2], Voutsinas M[9], Zane L[1,2]*

[1]Department of Biology, University of Padova, via U. Bassi 58/B, 35131 Padova, Italy

[2]Consorzio Nazionale Interuniversitario per le Scienze del Mare (CoNISMa), Piazzale Flaminio 9, 00196 Roma, Italy

[3]Institut National Agronomique de Tunisie (INAT), 43 Avenue Charles Nicolle 1082, Tunis, Mahrajène Tunisie

[4]USR 3278 CNRS-EPHE, CRIOBE, Université de Perpignan Via Dominitia, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France

[5]Department of Biological and Environmental Sciences and Technologies, University of Salento, Strada Provinciale Lecce-Monteroni, 73100 Lecce, Italy

[6]University of Zadar, Ul. Mihovila Pavlinovića, 23000 Zadar, Croatia

[7]Institute of Marine Biology Kotor (IBMK), Dobrota bb, PO box 69, 85330 Kotor, Montenegro

[8]National Institute for Marine Research and Development "Grigore Antipa", 900581 Constanţa, Romania

[9]Hellenic Centre for Marine Research, 46.7 km Athens Souniou Ave., PO box 712, 19013 Anavyssos, Greece

Keywords: Mediterranean mussel, population genomics, 2b-RAD, SNP markers, marine protected areas, MPAs

Running title: *Mytilus galloprovincialis* in the Mediterranean and Black Seas

*Corresponding author: Lorenzo Zane
Department of Biology, University of Padova, Via U. Bassi 58/B, 35131 Padova, Italy
Phone: +39 0498276220, Fax: +39 0498276209, E-mail: lorenzo.zane@unipd.it

Introduction

Understanding the distribution of genetic variability is a keystone for environmental resources management and conservation biology of marine species (Moritz 1994; Palumbi *et al*. 2003; Cowen *et al*. 2006). Population connectivity plays a crucial role in local and meta-population dynamics, genetic structure and population resiliency to human exploitation (Hastings and Harrison, 1994; Cowen et al. 2007; Weersting and Toonen 2009; Puckett et al. 2014). The majority of marine species has a planktonic larval stage that, dispersing over days up months in the currents, is supposed to be the primary source of the dispersal capability (Mileikovsky, 1971; Ward et al. 1994; Gilg and Hilbish, 2003). In this framework, especially when the working conditions make the direct measure of dispersal of species difficult, such as the tracking of larvae, genetic data are widely used for the inference of population connectivity (Cowen and Sponaugle 2009; Hellberg et al. 2002; Palumbi 2003; Thorrold et al. 2002; Broquet and Petit 2009; Lowe et al. 2010).

The FP7 CoCoNET European project aimed at an in-deep analysis of connectivity and of those parameters involved in the transport of propagules among present and future marine protected areas (MPAs) in the Mediterranean and Black seas. The main task is the exploration of the population structure of selected species from different taxonomic groups with different ecological roles and dispersal mechanisms. Specific traits related to the biology of each species, such as the pelagic larval duration (PLD), the reproductive timing, behavioral features or the recruitment success, can result in very different connectivity outcomes (Boissin et al. 2016; Jahnke et al. 2016; Carreras et al. 2017; Paterno et al. 2017; Jahnke et al. submitted, Marino et al. submitted), thus multispecies genetic analysis is needed to obtain a "collection of networks of genetic variation of all species within a community" (Fortuna et al. 2009).

The Mediterranean mussel *Mytilus galloprovincialis* (Lamarck, 1819) was selected as one of the target species in the CoCoNET framework for the assessment of genetic connectivity in the Mediterranean (within a pilot area in Adriatic-Ionian basins) and in the Black Sea. *M. galloprovicincialis* is an economically important species (Astorga et al. 2014) widely used in environmental bio-monitoring for the sessile life style and the filter-feeding habit (Gosling 1992). Generally, marine mussels of genus *Mytilus* occur worldwide in all oceans and major seas in both northern and southern hemispheres (Gerard et al. 2008; Kijewski et al. 2011) showing a wide distribution range. Specifically, in Europe the presence of 3 distinct *Mytilus* taxa (*M. galloprovincialis*; *M. edulis* Linnaeus, 1758; *M. trossulus* Gould, 1850) and their hybrids (*M. edulis*/*M. trossulus* and *M. edulis*/*M. galloprovicialis*) has been traditionally reported. *M. galloprovincialis* originated in the Mediterranean Sea around 2 million years ago (Daguin and Borsa 2000) and, at present, it occurs as a pure taxon in the latter and in the Black Sea (Smietanka et al. 2004, Kijewski et al. 2011). So far, few genetic studies describing the genetic structuring of this species along the Mediterranean and Black seas' coasts have been conducted, but they are far away from being comprehensive because of the restricted number

of sampling sites, the low population samples size, and the limits of the molecular markers themselves. In fact, high resolution, hyper variable nuclear markers such as microsatellites show typically low densities and difficulty of isolation in molluscs (Cruz et al. 2005; McInerney et al. 2011), whereas the use of mitochondrial markers is complicated by the doubly uniparental inheritance (DUI) of these organelles in this taxon (Skibinski et al. 1994; Zourus et al. 1994). At a wide geographic scale, a restricted gene flow across Atlantic, Mediterranean and Black Seas is suggested (allozyme loci and RFLPs; Sanjuan et al. 1996; Quesada et al. 1995; Sanjuan et al. 1997, Ladoukakisn et al. 2002), whereas, at a smaller scale, genetic homogeneity was found within the Aegean Sea and in the Central-Eastern Mediterranean Sea, and genetic heterogeneity was evident in the Adriatic-Ionian basins (microsatellites, mtDNA F and M; Ladoukakisn et al. 2002; Giantsis et al. 2014a, 2014b). If in the Adriatic-Ionian area the genetic data about this species are scarce because of the very limited number of localities considered (up to 3; Giantsis et al. 2014a, 2014b), in the Black Sea its genetic state is essentially unknown.

In the present study, we applied a population genomics approach based on hundreds polymorphic loci and we took advantage of a wide sampling design consisting of 8 localities in the Adriatic-Ionian basins, 2 in the Western Mediterranean and 9 localities in the Black Sea to providing the first insight into the *M. galloprovincialis* genetic structure in these areas. We used the RADseq 2b-RAD protocol (Wang et al. 2012) for the isolation and genotyping of single nucleotide polymorphisms (SNPS) with a genome-wide coverage.

Specifically, we aimed to: 1) provide the overall genetic structure of *M. galloprovincialis* at the large scale analysing the population samples from the Western and Central Mediterranean and the Black seas; 2) investigate the presence of genetic differentiation at a small scale in the Central Mediterranean (Adriatic-Ionian basins); 3) investigate the genetic differentiation in the Black Sea; 4) compare our findings to those obtained for others CoCoNET species in the same study area; and 5) discuss the outcomes in the light of the design and management of present and future MPAs.

Methods

Sampling and genomic DNA extraction

478 individuals of *Mytilus galloprovincialis* were collected from 2013 to 2014 in the Mediterranean and Black seas at 19 sampling locations (9 sites in Black Sea, 8 sites in Adriatic-Ionian seas and 2 site in the Western Mediterranean Sea; see Table 1 and Fig. 1 for details). At each location, the shell of each sample was opened, the body removed and preserved in absolute ethanol until the genomic DNA (gDNA) extraction from gills (Eurogold Tissue DNA mini kit, EuroClone). The quality of extracted gDNA in terms of fragments size (1% agarose gel electrophoresis), concentration and purity ratio (NanoDrop UV-Vis spectrophotometer) was checked for all the samples.

Construction and Next Generation Sequencing of the 2b-RAD libraries

The concentration of high-quality RNA-free gDNA obtained from *M. galloprovincialis* samples was optimized for the 2b-RAD protocol (Wang et al. 2012) at about 250-400 ng for each individual. All the collected samples (N=278) were processed following the 2b-RAD protocol steps as described in Paterno et al. (2017), except for i) the ligation of adaptors with partially (not fully) degenerated overhangs to the restriction fragments and ii) the 2b-RAD tags amplification for a smaller number of cycles (14). Three pools (about 160 barcoded-samples each) were assembled and the high-molecular weight fragments and primer-dimers removed in two steps as in Paterno et al. (2017). The Next Generation Sequencing of each pool was performed on Illumina HiSeq platforms with a single-end SR50 High Output mode by UC Davis Genome Center (CA, USA) and Genomix4Life S.r.l. (Baronissi, SA, Italy); the sequencing services performed also the data demultiplexing and quality-filtering. For increasing the read depth and ensuring an equal number of reads for each sample, each pool was sequenced twice (Paterno *et al.* 2017).

***De novo* analysis: *In silico* identification of loci and genotyping**

The quality of the raw demultiplexed and quality-filtered reads was verified by using FASTQC 0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Later, custom-made Python™ scripts were used for trimming the adaptors and filtering the reads for the presence of the IIB restriction enzyme (*CspCI*), producing trimmed high-quality reads of uniform 32bp-length.

STACKS 1.42 (Catchen et al. 2013) was used to carry out the assembly of loci and the individual genotyping (*denovo_map.pl* pipeline. Two technical replicates of 2 samples (one from the Mediterranean Sea and another one from the Black Sea) were included for the optimization of the *de-novo* assembly parameters in STACKS. Parameters were set as follows: *m*=15 (minimum stack depth per allele in a sample), *M*=4 (number of mismatches allowed between stacks to build a locus in an individual), *n*=4 (maximum distance between loci from distinct individuals to be merged in the population catalogue), *SNP bound*=0-0.1 (error rate to call SNP), *alpha*=0.05 (significance level to call a heterozygote or homozygote).

The employment of SNPs as RAD-Seq markers in population genomics typically implicates a certain level of missing data in the dataset because not all the genotyped markers are shared by all the samples (Chattopadhyay et al. 2014), which implies the need to retain only the loci shared among a reasonable, yet arbitrary, fraction of individuals for downstream analysis. However, if the missing data are not evenly scattered among the samples, a dataset of many individuals collected from distinct areas could lose power in detecting a fine genetic differentiation within areas because of the decrease of the number of shared polymorphic markers usable for the analyses.

For this reason, we performed 3 independent STACKS runs in order to achieve the best power of resolution according to the geographical scale considered: the large scale (the entire study area) on one side, and the basin scale (Adriatic-Ionian basins or Black Sea) on the other side. First, an overall catalogue of loci for the exploration of the global population structure of the species throughout the entire study area was obtained

by running together the 478 samples (plus the 2 technical replicates) from all 19 population samples. Later, the 207 samples (plus one technical replicate) from the 8 Adriatic-Ionian sites and the 232 samples (plus one technical replicate) from the 9 Black Sea's sites were run separately to obtaining a basin-specific dataset for the evaluation of the genetic structuring of the species at a small spatial scale.

For each of the 3 STACKS runs, we selected a final dataset consisting of all the polymorphic loci shared by at least the 80% of the individuals with up to 3 SNPs and 6 alleles. When more than one SNP was found at a locus, only the SNP with the highest expected heterozygosity across each dataset was retained (Phillips 2005). The threshold of missing loci per individual was set to 30%. Since the parameter *m* (minimum read depth for allele) was set to 15, no additional filter for sequencing depth was applied (Paterno et al. 2017). At the end, we obtained 3 distinct datasets: 1) the overall dataset, 2) the Adriatic-Ionian dataset and 3) the Black Sea's dataset. The package *Populations* in STACKS was used to generate input file in GENEPOP format, and later CREATE (Coombs et al. 2008) and PGDSPIDER 2.1.0.3 (Lischer and Excoffier 2012) for the conversion in several formats for genetic analyses.

Statistical power of the basin-specific datasets

The statistical power to detect true levels of genetic differentiation was evaluated and compared for the Adriatic-Ionian and Black Sea's datasets by the simulation method of Ryman and Palm (2006) implemented in POWSIM 4.1 (SNPs executable Powsim_b). POWSIM simulates sampling from populations at various levels of expected divergence simulating a classical Wright-Fisher model without migration or mutation. Simulations were run using default parameter values for dememorizations (1000), batches (100), and iterations per batch (1000) for a scenario involving 8 and 9 subpopulations, respectively, using a range of effective population size (N_e=1000-10 000) and of generations of drift (t=10-100). The statistical power to detect an expected divergence as small as F_{ST}=0.005 was estimated after 200 replicates as the proportion of statistically significant test ($p<0.05$).

Population genomics

The genetic variability within population samples, population structure and genetic differentiation were evaluated separately for the 3 datasets (overall, Adriatic-Ionian and Black Sea's datasets) in order to achieve an appropriate power of resolution to variable geographical scales. The observed (H_o) and unbiased expected heterozygosity (H_e) of each population sample were assessed on the polymorphic loci (GENETIX 4.05.2; Belkhir et al. 2000) and the allelic richness (A_R) was calculated on the smallest population sample size of each dataset (HP-RARE; Kalinowski 2005). A non-hierarchical and hierarchical analysis of molecular variance (AMOVA; Excoffier et al. 1992) as well as the pairwise genetic distances between population samples (F_{ST}) were calculated (ARLEQUIN 3.5; Excoffier and Lischer 2010); due to the presence of missing data, the locus by locus option was set. Benjamini and Hochberg (1995) correction for multiple tests was applied for adjusting the significance level in multiple comparisons. The Discriminant Analysis of Principal

Components (DAPC; ADEGENET for R Jombart et al. 2010) was used to find the best number of clusters of genetically related individuals (based on the Bayesian Information Criterion, BIC) and for the visual assessment of the between-population differentiation of the original population samples. Bayesian clustering (STRUCTURE 2.3.4; Pritchard et al. 2000) was performed to investigating the most likely number of genetic clusters. After some trials, a burn-in of 100 000 followed by one million runs for K=1-8 (depending on the datasets) and 5 iterations per each K value was set; sampling locations as prior (model LOCPRIOR, Hubisz et al. 2009) was used; admixture model and correlated allelic frequencies were assumed. The most probable number of clusters was identified on delta K (STRUCTURE HARVESTER online; Earl and Vonholdt 2012).

Detection of loci under selection (outliers)

For the identification of putative loci under directional selection, two neutrality tests were employed. The first test is the Bayesian method implemented in BAYESCAN 2.01 (Foll and Gaggiotti 2008), and the second one is the F_{ST}-outlier method implemented in LOSITAN (Antao et al. 2008). Both softwares were run separately with the 3 datasets. LOSITAN test was set as follows: 1 million simulations under neutral mean F_{ST}, confidence interval of 0.95% and a false discovery rate (FDR) of 0.01; the infinite allele model was applied. For each run, three replicates were performed. BAYESCAN run with burn in=50 000, thinning interval=30, sample size=5000, numbers of pilot runs=50, length of pilot runs=5000 and the same false discovery rate (FDR) threshold set in LOSITAN (0.01). All the loci detected by these 2 methods were considered outliers under directional selection.

Results

Sequencing results, filtering and selection of loci for genetic analysis

For all collected *M. galloprovincialis* samples (N=478), high-quality RNA-free gDNA was successfully extracted from gills. The Illumina sequencing of the three 2b-RAD pools (N=480; 2 technical replicates included) produced 1 032 760 048 de-multiplexed and filtered-quality reads (mean quality score per base higher than 37) and more than 92% was retained after trimming and filtering for *CspCI* restriction site (about 1 990 000 reads for each sample). The 3 independent STACKS runs identified 113 494, 77 459 and 83 000 2b-RAD tags among the total, the 8 Adriatic-Ionian and 9 Black Sea's population samples, respectively; about the 33% of loci were polymorphic in all 3 runs.

After filtering and the exclusion of the 2 technical replicates and of those individuals showing more than 30% of missing loci, we obtained 3 datasets typified as follows: 1) the overall dataset: 461 individuals, 512 polymorphic loci; 2) the Adriatic-Ionian dataset: 201 individuals, 811 polymorphic loci; and 3) the Black Sea's dataset: 228 individuals, 998 polymorphic loci.

Genetic analyses

Statistical power of the basin-specific datasets

The 200 simulations performed by POWSIM showed that the population sample size, number of loci and their allele frequencies of the 2 basin-specific datasets have enough statistical power to detect F_{ST} value=0.005 with a probability approaching the 100% under all the tested conditions of effective population sizes and number of generations of drift.

Genetic variability

For the 3 datasets, similar values were obtained among the population samples within each one for H_o, unbiased H_e and A_R, confirmed by one-way ANOVA detecting no significant differences among samples (P>0.05; Tables S1, S2 and S3; Supporting Information).

Genetic structure at the large scale

The non-hierarchical AMOVA carried out on the 19 population samples (the overall dataset, 512 polymorphic loci) provided a strong indication of genetic structuring at the large spatial scale, producing a global multilocus F_{ST} statistically significant (F_{ST}=0.03243, P<0.0001). Considering the 171 pairwise F_{ST} (Table 2), all the 90 comparisons involving a location from the Mediterranean Sea and a location from the Black Sea were statistically significant (F_{ST} range=0.04208-0.08079; P<0.0001, significant after Benjamini and Hochberg correction for multiple tests), indicating a clear genetic differentiation between the 2 seas. In addition, 14 comparisons were significant among the 45 tests between Mediterranean samples (F_{ST} range=0.00809-0.02075; P<0.05), due mainly to the deviation of the sample from Montenegro (BOK, 9 significant comaprisons). The North-Western (France) and the South-Western (Tunisia) samples were not differentiated each from the other despite the geographic distance, and they differed only weakly from the Central Mediterranean ones (3 significant comparisons, 2 of them involving the BOK sample). At last, no significant pairwise F_{ST} was detected in the 36 comparisons within the Black Sea.

Accordingly, a hierarchical AMOVA showed that the genetic variation of the dataset can be best partitioned into 2 groups, one including the 10 Mediterranean population samples, and the other one the 9 Black Sea's populations samples (F_{ST}=0.05286, P<0.0001) maximizing the genetic variance between groups. Similarly, the DAPC identified two genetic clusters as the optimal clustering solution based on the lowest BIC (Fig. 2A), and these 2 clusters fitted to the Mediterranean and the Blacks seas' population samples (Figs. 2B and 2C). Interestingly, the 2D graphs plotting the density of the inferred clusters and the original populations on the single discriminant function described the same trend (Figs. 2C and 2D). The Bayesian clustering carried out by STRUCTURE was consistent with the DAPC, in fact the most probable number of clusters was identified for K=2 with the same groupings (Fig. 3).

Genetic structure at the basin scale

The lack of genetic structure of the Black Sea and the signal of genetic differentiation in the Central Mediterranean found with the overall dataset, were confirmed by the basin-specific datasets.

The non-hierarchical AMOVA conducted on Adriatic-Ionian dataset (811 polymorphic loci) pointed out that most of the genetic variation (>99%) arose from the within-populations level and only 0.69% from the among-populations level, with a relative global F_{ST} statistically significant (P<0.0001). A number of pairwise F_{ST} values were statistically significant after correction for multi-test (9/28; Table 3) and, once again, most of these comparisons (6/9) involved the Montenegro's sample (BOK); 2 pairwise comparisons involved the Croatia's sample (KOR) and the Tremiti islands (TRE) and Otranto (OTR), and the last one between TRE and OTR. Both the genetic clustering methods found out a single cluster including the 8 population samples (BIC values of the DAPC in Fig.4; STRUCTURE plot not provided). Interestingly, a hierarchical AMOVA found the best clustering solution for 3 groups: the first group with the population sample from Tremiti islands (TRE) alone, the second group including BOK sample along with the Croatian sample (KOR), and at last the 4 Italian samples together with the Greek (OTH-A) and Albanian (KAP) samples. This configuration highlighted an East-West differentiation in the Central Adriatic Sea, and a general homogeneity in the Southern Adriatic and Ionian Sea. The scatterplot of the discriminant analysis set on the original populations reflected the weak genetic differentiation detected among these population samples (Fig. 4B and 4C).

The Black sea's dataset (998 polymorphic loci) found out a small and not significant global multilocus F_{ST} value (F_{ST}=0.00278, P=0.59584) and no one pairwise F_{ST} value resulted significant within this basin. Both Bayesian clustering methods conducted by ADEGENET and STRUCTURE found only one genetic cluster (BIC values of the DAPC in Fig. 5; STRUCTURE plot not reported). As expected because of the lack of genetic structuring, the visual assessment of the between-population differentiation of the original populations showed the 9 samples overlapping and not spread out (DAPC plots, Figs. 5B and 5C).

Outlier detection

The outliers detection conducted on the overall dataset identified a total of 25 loci putatively under positive selection, 14 of these were found out by both softwares and 11 only by BAYESCAN. Although the exclusion of these loci resulted in a substantial reduction of the global genetic variance of the dataset (F_{ST}=0.01152; P<0.0001), the general pattern of structuring was confirmed, with the best clustering solution for 2 groups (Mediterranean population samples against Black Sea's population samples; F_{CT}=0.01565; P<0.0001) and with 86/90 significant pairwise F_{ST} distances between the Mediterranean Black seas.

The scan of the Adriatic-Ionian dataset detected 2 putative outliers (identified by both softwares), whose exclusion caused a small decrease of the molecular variance (F_{ST}=0.00591; P<0.0001) and the loss of 3 significant pairwise F_{ST}, all involving the population sample from Montenegro (BOK).

As expected due to the complete lack of genetic differentiation, the scan of the Black Sea's dataset did not identify any locus putative under selection within the Black Sea.

Discussion

The genomic scan conducted on the 19 *Mytilus galloprovincialis* population samples from the Mediterranean and Black Seas, analysed at different spatial scale, provided the first detailed analysis of the population structure of this species throughout its native range. In this study we used the 2b-RAD protocol (Wang et al. 2012) for the identification and genotyping of hundreds SNPs with a genome-wide coverage and specifically we used 3 different datasets to achieve the best power of resolution for the large scale (the entire study area) and the basin scale (Adriatic-Ionian basins or Black Sea).

Genetic structure at the large scale

A marked structuring between the Mediterranean and Blacks seas was found, revealing that these 2 seas are genetically distinct. These results are not surprising and they match with previous genetic data obtained for this species, that detected a restrict gene flow between the Mediterranean and Black Sea (Ladoukakisn et al. 2002). Such differentiation was previously reported also in others species (e. g. stickelbacks; Makinen and Merila 2008) and was explained by the effect of oceanographic features, like the presence of Bosphorus and Marmara Sea, and the existence of natural barriers (salinity, temperature) preventing the exchange between the 2 seas. The evaluation of the genetic variability distribution at wide scale wasn't effected by the presence of a few loci putative under selection. In fact, although their exclusion caused a reduction of the overall genetic variance, the pattern of the genetic structure remained consistent, with a clear genetic break weetween the Mediterranean and the Black seas.

Genetic structure at the basin scale

Within the Mediterranean Sea, no signal of differentiation was found between the Western and Central, as it was expected considering results obtained for other species in previous studies (Borsa et al. 1997, Paterno et al. 2017). Contrary, the Adriatic-Ionian population samples showed a weak genetic differentiation, mostly in pairwise comparisons involving the population from Montenegro (BOK). The Bayesian clustering failed in detecting genetically differentiated clusters (both methods employed found only one cluster) likely due to the weakness of the differentiation. On the other hand, a hierarchical AMOVA found that genetic variation of the dataset can be best partitioned into 3 groups: a first group including the Tremiti islands's sample (TRE) alone, the second one comprising Montenegro's sample (BOK) along with the Croatian one (KOR), and at last the 4 Italian samples together with the Greek (OTH-A) and Albanian (KAP) samples. This differentiation pattern underlined an East-West differentiation in the Central Adriatic Sea, whereas a general homogeneity was found in the Southern Adriatic and Ionian seas among the Italian, Greek and Albanian localities. Being the adult sessile, the dispersal ability in this species and thus the gene flow is mostly due to the larval stage, that is likely constrained by physical and ecological factors like the complex water circulation of the Adriatic Sea, characterized by the presence of minor gyres in addition to the main current. The gyral circulation could prevent the exchange of larvae, acting as barriers

among localities between East and West Central Adriatic coasts, also determing a certain level of self retention in others. These genetic results are consistent with findings of Giantsis and colleagues obtained with mtDNA and microsatellites (2014a; 2014b), who found genetic heterogeneity among Adriatic samples, even if their sampling effort was very limited (2-3 sampling sites). The genetic variability distribution at a local scale in the Central Mediterranean wasn't effected by the presence of a few outliers putatively under selection identified by 2 neutrality tests. Their exclusion from the analisys caused a sostantial reduction of the genetic variance of this dataset, but the pattern of the genetic differentiation remained consistent with an East-West differentiation in the Central Adriatic Sea and genetic homogeneity in the Southern Adriatic and Ionian seas.

Moving the attention to the Black Sea, a completely lack of genetic structure was detected. In this basin, the currents pattern and the long pelagic larval duration probably ensure a good connection among the sampled locations. It is not possible to compare these data with previous results because these findings represent the first information regarding *M. galloprovincialis* population structure in this sea.

Comparison with other CoCoNET species

The genetic findings of this research are similar to the pattern of differentiation obtained in other CoCoNET species. The marble crab *Pachygrapsus marmoratus* population samples collected at the same Adriatic-Ionian localitites showed a fairly genetic homogeneity, even if a weak signal of differentiation was evident after removing a locus putatively under selection (Marino et al. under revision). This data was explained with the existence of random fluctuations of allele frequencies depending on variability in fecundity, mortality or reproductive success, that can lead to a scenario of genetic patchiness in a general context of genetic homogeneity, yet previously reported for this species (Silva et al. 2009; Fratini et al. 2011, 2013). This genetic data was fully in line with the outputs of Lagrangian simulations, pointing out a high level of propagules exchange among sites, although the presence of seasonal and interannual variation in the oceanographic connectivity intensity was detected (Marino et al. under revision).

The black scorpionfish *Scorpaena porcus* is a benthic species that dispersing mainly during the pelagic larval stage. This species showed a clear genetic break at the wide scale between the Mediterranean and Black Seas, and an East-West differentiation in the Central Mediterranean within the Adriatic-Ionian basins seemed to prevail, with a clear genetic break between the eastern and western localities (Boissin et al. 2016). In the Black Sea a general homogeneity was observed, with a weak signal of differentiation only for 2 isolated localities. The authors didn't identify a clear geographic pattern for this differentiation, that was attributed to chaotic genetic patchiness (Boissin et al. 2016).

The east Atlantic peacock wrasse *Symphodus tinca* (Carreras et al. 2017) is a demersal fish supposing to disperse only during the larval stage. The evaluation of the population structure of this species in the Adriatic-Ionian basins identified the existence of 2 barriers to gene flow separating the Eastern shore

localities from the South-Western ones and differentiating the Tremiti islands's sample (TRE) from all the others (Carreras et al. 2017).

At the extremities of this differentiation range, we can find the sea urchin *Paracentrotus lividus* (Paterno et al. 2017) and the habitat former *Posidionia oceanica* (Jahke et al. in review). The analysis of *P. lividus* samples indicated a clear genetic homogeneity within and between Adriatic-Ionian basins, with a broadly consistent prediction of the potential larval exchange among the same sites (obtained with Lagrangian simulations) that supports the view of a relatively persistent flux of propagules from each sites to the closest neighbor, mostly in a counterclockwise direction (Paterno et al. 2017). On the other side, the seagrass *P. oceanica* showed a marked genetic structuring in the same area, even between the closest localities, with a possible north-to-south subdivision (Jahke et al. in review).

Implications for MPAs

MPAs have been increasingly considered the most effective tool for conservation and management of marine environment, able to reduce the alteration of ecosystems and mitigating the negative effect of human activites (Gaines et al. 2010; Gabrié C. et al. 2012). The implementation of networks of MPAs, where the protected areas can cooperatively operate, is the new conservation target, because it's supposed that networks could be more effective and offer more protection than the individual sites alone (Wood et al. 2008; Wabnitz et al. 2010; Fenberg et al. 2012; Grorud-Colvert et al. 2014). Connectivity data obtained so far in the CoCoNET study area in the central Mediterranean and Black Sea, highlighted species-specific connectivity patterns, ranging from the total absence of genetic differentiation (*P. lividus*) to a strong structuring (*P. oceanica*). All the target species have larval dispersal but a different adult life style. Species-specific life history traits (PLD, reproductive timing) and the interactions with environmental settings and ecological factors (temperature, oceanographic barriers) can determine the different connectivity patterns observed. These studies underlined the importance of evaluating and discussing the genetic data of population structure in a seascape framework, thus in the light of those biological, physical and ecological factors that could have a significant role in shaping the genetic variability of these species. In a management and conservation perspective, the implementation of ecological effective networks of MPAs should rely on the knowledge of connectivity patterns of a representative panel of species living in the selected areas. The heterogeneity of the connectivity outcomes so far obtained i) indicated the importance of evaluate the genetic distribution of the species under a seascape genomics framework (Di Franco et al. 2012; Pujolar et al. 2013; Paterno et al. 2017); ii) warns against any generalization in defining conservation units based on the results of the single species because not representative of the entire community (Melià et al. 2016); iii) suggested to take into account connectivity data of umbrella species with a variety of life history traits to improve the efficacy of the management conservation strategy of the marine ecosystem (Pascual et al. 2017).

Acknowledgements

The research was supported by the European Community's 7th Framework Programme (FP7/2007−2013) under Grant Agreement No. 287844 for the project "Towards COast to COast NETworks of marine protected areas (from the shore to the high and deep sea), coupled with sea-based wind energy potential (CoCoNet, http://www.coconet-fp7.eu/)", and by the University of Padua grant CPDA148387/14.

Author contributions

MP and LZ conceived and designed he experiments and wrote the manuscript. MP performed the experiments. MP and LZ analyzed the data. LZ contributed reagents/materials/analysis tools. MP and LZ wrote the manuscript that was reiewed and approved by all the co-authors.

List of brief titles of items in the supplementary material

Table S1. Genetic variability at 512 polymorphic loci (overall dataset) of 19 *Mytilus galloprovincialis* population samples from Mediterranean and Black seas.

Table S2. Genetic variability at 811 polymorphic loci (Adriatic-Ionian dataset) of 8 *Mytilus galloprovincialis* populations samples from Adriatic and Ionian seas.

Table S3. Genetic variability at 998 polymorphic loci (Black Sea's dataset) of 9 *Mytilus gallorovincialis* populations samples from Black Sea.

References

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A Workbench to Detect Molecular Adaptation Based on a F_{ST}-Outlier Method. *BMC Bioinformatics*, 9, 323.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2000) GENETIX 4.02, Logiciel Sous Windows TM Pour La Génétique Des Populations. Laboratoire Génome, Populations Interactions, CNRS UMR 5000.

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, 289–300.

Boissin E, Micu D, Janczyszyn-Le Goff M, Neglia V, Bat L, Todorova V, Panayotova M et al. (2016) Contemporary Genetic Structure and Post-Glacial Demographic History of the Black Scorpionfish, *Scorpaena Porcus*, in the Mediterranean and the Black Seas. *Molecular Ecology*, 25, 2195–2209.

Broquet T, Petit EJ (2009) Molecular Estimation of Dispersal for Ecology and Population Genetics. *Annual Review of Ecology, Evolution and Systematics*, 40, 193–216.

Carreras C, Ordóñez V, Zane L, Kruschel C, Nasto I, Macpherson E, Pascual M (2017) Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation. *Scientific Reports*, 7:43417.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: An Analysis Tool Set for Population Genomics. *Molecular Ecology*, 22, 3124–3140.

Chattopadhyay B, Garg KM, Ramakrishnan U (2014) Effect of diversity and missing data on genetic assignment with RAD-Seq markers. *BMC Research Notes*, 7:841.

Coombs JA, Letcher BH, Nislow KH (2008) Create: a Software to Create Input Files from Diploid Genotypic Data for 52 Genetic Software Programs. *Molecular Ecology Resources*, 8, 578–580.

Cowen RK, Paris CB, Srinivasan A (2006) Scaling of Connectivity in Marine Populations. Science, 311, 522–27.

Cowen RK, Gawarkiewicz G, Pineda J, Thorrold SR, Werner FE (2007) Population connectivity in marine systems: An overview. *Marine Biology and Fisheries*, 20, 14–21.

Cowen RK and Sponaugle S (2009) Larval Dispersal and Marine Population Connectivity. *Annual Review of Marine Science*, 1, 443–66.

Cruz F, Montse P, Presa P (2005) Distribution and Abundance of Microsatellites in the Genome of Bivalves. *Gene*, 346, 241–47.

Daguin C, Borsa P (2000) Genetic relationships of *Mytilus galloprovincialis* populations worldwide: evidence from nuclear-DNA markers. In Harper EM, Taylor JD, Crame JA eds. The evolutionary biology of the Bivalvia. *Geological Society of London Special Publication*, 177, 389–397.

Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361.

Excoffier L, Hofer T, Foll M (2009) Detecting Loci Under Selection in a Hierarchically Structured Population. *Heredity*, 103, 285–298.

Excoffier L, Lischer HEL (2010) Arlequin Suite ver 3.5: a New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564–567.

Foll M and Gaggiotti OE (2008) A Genome Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics, 180, 977–993.

Fortuna MA, Albaladejo RG, Fernández L, Aparicio A, Bascompte J (2009) Networks of spatial genetic variation across species. *PNAS*, 106, 19044–19040.

Gerard K, Bierne N, Borsa P, Chenuil A, Feral JP (2008) Pleistocene separation of mitochondrial lineages of *Mytilus* spp. mussels from Northern and Southern Hemispheres and strong genetic differentiation among southern populations. *Molecular Phylogenetics and Evolution*, 49, 84–91.

Giantsis IA, Mucci N, Randi E et al. (2014a) Microsatellite variation of mussels (*Mytilus galloprovincialis*) in central and eastern Mediterranean: genetic panmixia in the Aegean and the Ionian Seas. *Journal of the Marine Biological Association of the UK*, 94, 797–809

Giantsis IA, Abatzopoulos TJ, Angelidis P, Apostolidis AP (2014b) Mitochondrial Control Region Variability in *Mytilus galloprovincialis* Populations from the Central-Eastern Mediterranean Sea. *International Journal of Molecular Sciences*, 15, 11614–11625.

Gilg MR, Hilbish TJ (2003) The geography of marine larval dispersal: coupling genetics with fine-scale physical oceanography. *Ecology*, 84, 2989–2998.

Gosling E (1992) Genetics of *Mytilus*. In: Gosling, E. (Ed.) The Mussels *Mytilus*: Ecology, Physiology, Genetics and Culture. Elsevier, The Netherlands, pp. 309–382.

Hastings A, Harrison S (1994) Metapopulation Dynamics and Genetics. *Annual Review of Ecology and Systematics*, 25:167–188.

Hellberg ME, Burton RS, Neigel JE and Palumbi SR (2002) Genetic Assessment of Connectivity among Marine Populations. *Bulletin of Marine Science*, 70, 273–90.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9, 1322–1332.

Jahnke M, Christensen A, Micu D, Milchakova N, Sezgin M, Todorova V, Strungaru S, Procaccini G (2016) Patterns and mechanisms of dispersal in a keystone seagrass species. *Marine Environmental Research*, 117, 54–62.

Jahnke M, Casagrandi R, Melià P, Schiavina M, Schultz S, Zane L, Procaccini G (in review, *Diversity & Distribution*). Connectivity, Sexual Reproduction and Selection Pressure in the Seagrass *P. oceanica* and the Implications for Conservation.

Jombart T (2008) Adegenet: A R Package for the Multivariate Analysis of Genetic Markers. *Bioinformatics*, 24, 1403–1405.

Kalinowski ST (2005) HP-RARE 1.0: A Computer Program for Performing Rarefaction on Measures of Allelic Richness. *Molecular Ecology Notes*, 5, 187–189.

Kijewski T, Śmietanka B, Zbawicka M, Gosling E, Hummel H, Wenne R (2011) Distribution of *Mytilus* taxa in European coastal areas as inferred from molecular markers. *Journal of Sea Research*, 65, 224–234.

Ladoukakis ED, Saavedra C, Magoulas A, Zouros E (2002) Mitochondrial DNA Variation in a Species with Two Mitochondrial Genomes: The Case of *Mytilus galloprovincialis* from the Atlantic, the Mediterranean and the Black Sea. *Molecular Ecology*, 11, 755–69.

Lischer HEL, Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28, 298–299.

Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, 19, 3038–3051.

Marino IAMM, Schiavina M, Aglieri G, Boscari E, Congiu L, Faggion S … Melià P (in review, *Molecular Ecology*) Assessing Connectivity Patterns of the Marbled Crab *Pachygrapsus marmoratus* in the Adriatic and Ionian Seas Combining Genetic Data and Lagrangian Simulations.

McInerney CE, Allcock AL, Johnson MP, Bailie DA, Prodohl PA (2011) Comparative Genomic Analysis Reveals Species-Dependent Complexities That Explain Difficulties with Microsatellite Marker Development in Molluscs. *Heredity*, 106, 78–87.

Mileikovsky SA (1971) Types of Larval Development in Marine Bottom Invertebrates, Their Distribution and Ecological Significance: A Re-Evaluation. *Marine Biology*, 10, 193–213.

Moritz C (1994) Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecololgy & Evolution*, 9, 373–5.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, 13, 146–58.

Pascual M, Rives B, Schunter C, Macpherson E (2017) Impact of life history traits on gene flow: A multispecies systematic review across oceanographisc barriers in the Mediterranean Sea. *PLos ONE*, 12, e0176419.

Paterno M, Schiavina M, Aglieri G, Ben Souissi J, Boscari E, Casagrandi R, Chassanite A, Chiantore M, Congiu L, Guarnieri G, Kruschel C, Macic V, Marino IAMM, Papetti C, Patarnello T, Zane L, Melià P (2017) Population genomics meet Lagrangian simulations: oceanographic patterns and long larval duration ensure connectivity among *Paracentrotus lividus* populations in the Adriatic and Ionian seas. *Ecology & Evolution*, 00:1–17.

Phillips C (2005) Using Online Databases for Developing SNP Markers of Forensic Interest. In A. Carracedo (Ed.), Forensic DNA Typing Protocols (pp. 83–105). Totowa, NJ: Humana Press.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155, 945–959.

Puckett BJ, Eggleston DB, Kerr PC and Luettich RA Jr (2014) Larval Dispersal and Population Connectivity among a Network of Marine Reserves. *Fisheries Oceanography*, 23, 342–61.

Quesada H, Beynon CM, Skibinski DO (1995) A Mitochondrial DNA Discontinuity in the Mussel *Mytilus galloprovincialis* Lmk: Pleistocene Vicariance Biogeography and Secondary Intergradation. *Molecular Biology and Evolution*, 12, 521–24.

Ryman N, Palm S (2006) POWSIM: A Computer Program for Assessing Statistical Power When Testing for Genetic Differentiation. *Molecular Ecology Notes*, 6, 600–602.

Sanjuan A, Comesaña AS, De Carlos A (1996) Macrogeographic Differentiation by mtDNA Restriction Site Analysis in the S.W. European *Mytilus galloprovincialis* Lmk. *Journal of Experimental Marine Biology and Ecology*, 198, 89–100.

Sanjuan A, Zapata C, Alvarez G (1997) Genetic Differentiation in *Mytilus galloprovincialis* Lmk. throughout the World. *Ophelia*, 47, 13–31.

Skibinski DOF, Gallagher C, Beynon CM (1994) Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. *Genetics*, 138, 801–809.

Śmietanka B, Zbawicka M, Wołowicz M, Wenne R (2004) Mitochondrial DNA lineages in the European populations of mussels *Mytilus*. *Marine Biology*, 146, 79–92.

Thorrold SR, Jones GP, Hellberg ME, Burton RS, Swearer SE, Neigel JE, Morgan SG and Warner RR (2002) Quantifying Larval Retention and Connectivity in Marine Populations with Artificial and Natural Markers. *Bulletin of Marine Science*, 70, 291–308.

Wang S, Meyer E, McKay JK, Matz MV (2012) 2b-RAD: A Simple and Flexible Method for Genome-Wide Genotyping. *Nature Methods*, 9, 808–10.

Ward RD, Woodwark M, Skibinski DOF (1994) A comparison of genetic diversity levels in marine, fresh-water, and anadromous fishes. *Journal of Fish Biology*, 44, 213–232.

Weersing K and Toonen RJ (2009) Population Genetics, Larval Dispersal, and Connectivity in Marine Systems. *Marine Ecology Progress Series*, 393, 1–12.

Zouros E, Oberhauser Ball A, Saavedra C, Freeman KR (1994) An Unusual Type of Mitochondrial DNA Inheritance in the Blue Mussel *Mytilus*. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 7463–67.

Tables

Table 1. Sampling information about the 19 *Mytilus galloprovincialis* population samples collected from Mediterranean and Blacks seas involved in this study. For each population samples, reported are: area, nation, sampling location, acronym, coordinates, date of collection and number of samples (N).

Area	Nation	Sampling location	Acronym	Coordinates N – E	Date	N
Ionian Sea	Greece	Corfù island	OTH-A	39°47.522' – 19°54.211'	November 2013	28
Ionian Sea	Albania	Karaburun Peninsula	KAP	40°23.568' – 19°19.498'	June 2013	18
Adriatic Sea	Montenegro	Boka Kotorska Bay	BOK	42°23.252' – 18°34.178'	June 2013	30
Adriatic Sea	Croatia	Kornati islands	KOR	43°47.535' – 15°16.889'	June 2013	25
Adriatic Sea	Italy	Tremiti islands	TRE	42°08.315' – 15°31.437'	June 2013	27
Adriatic Sea	Italy	Torre Guaceto MPA	TOG	40°42.999' – 17°48.003'	May 2013	21
Adriatic-Ionian Sea	Italy	Otranto	OTR	40°06.554' – 18°31.153'	May 2013	29
Ionian Sea	Italy	Porto Cesareo MPA	POC	40°11.715' – 17°55.077'	May 2013	29
Western Med. Sea	Tunisia	Haouaria	TUN	37.050440 – 10.967000	October 2014	22
Western Med. Sea	France	Banyuls	FRN	42.482290 – 3.1374160	October 2014	17
Black Sea	Ukraine	Karadag MPA	KAR	44°54.37' – 35°15.33'	July 2013	29
Black Sea	Ukraine	Tharankut	TAR	45°20.03' – 32°33.09'	July 2013	19
Black Sea	Romania	Costinesti Monastery	COS	43°55.534' – 28°38.442'	July 2013	28
Black Sea	Bulgaria	Cape Kaliakra	KAL	43°24.712'28 – °21.001'	May 2013	28
Black Sea	Bulgaria	Ropotamo-Kiten	ROK	42°11.706' – 27°50.163'	June 2013	27
Black Sea	Turkey	Sile	SIL	41°10.986' – 29°36.736'	July 2013	30
Black Sea	Turkey	Sinop	SIN	42°0.964' – 35°10.956'	May 2013	22
Black Sea	Georgia	Batumi	BAT	41°41.279' – 41°42.078'	July 2013	26
Black Sea	Russia	Novorossiisk-Gelendzik	RUS			23

Table 2. Pairwise genetic distances (F_{ST}) between the 19 population samples based on the overall dataset of 512 polymorphic loci. F_{ST} indices and p-values are reported below and above the diagonal, respectively; significant indices in bold. Benjamini & Hochberg correction for multiple tests was applied. See Table 1 for location acronyms.

	OTH-A	KAP	BOK	KOR	TRE	TOG	OTR	POC	TUN	FRN	KAR	TAR	COS	KAL	ROK	SIL	SIN	BAT	RUS
OTH-A		0.3317	**0.0011**	0.0363	0.3853	0.4441	0.3554	0.1260	0.2912	0.1631	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
KAP	0.00370		**0.0022**	**0.0141**	0.6835	0.1098	**0.0006**	0.0326	0.2389	0.2244	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
BOK	**0.01052**	**0.01336**		**0.0171**	**0.0180**	**0.0028**	**<0.0001**	**<0.0001**	**0.0112**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
KOR	0.00739	**0.01161**	**0.00816**		**0.0191**	0.6951	**0.0041**	0.0339	0.3798	**0.0088**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
TRE	0.00280	0.00076	**0.00809**	**0.00933**		0.7108	0.0957	0.2115	0.2599	0.6005	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
TOG	0.00263	0.00819	**0.01113**	0.00042	0.00057		0.3195	0.7797	0.9893	0.4229	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
OTR	0.00297	**0.01740**	**0.01778**	**0.01139**	0.00653	0.00420		0.4884	0.0662	0.0391	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
POC	0.00490	0.00925	**0.01793**	0.00758	0.00412	-0.00060	0.00197		0.5505	0.1634	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
TUN	0.00421	0.00645	**0.01003**	0.00379	0.00518	-0.00432	0.00833	0.00196		0.1336	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
FRN	0.00648	0.00698	**0.02075**	**0.01349**	0.00213	0.00427	0.01080	0.00630	0.00895		**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
KAR	**0.04532**	**0.04870**	**0.04821**	**0.05202**	**0.05206**	**0.06536**	**0.05679**	**0.06200**	**0.05029**	**0.06607**		0.2067	0.9196	0.7956	0.4711	0.2631	0.3381	0.4182	0.2252
TAR	**0.05050**	**0.05875**	**0.04982**	**0.05634**	**0.05890**	**0.07106**	**0.05826**	**0.06710**	**0.05445**	**0.07511**	0.00381		0.1642	0.2921	0.0819	0.1958	0.5709	0.4627	0.2647
COS	**0.05049**	**0.05369**	**0.04960**	**0.05279**	**0.05845**	**0.06885**	**0.06117**	**0.06788**	**0.05583**	**0.06765**	-0.00210	0.00443		0.6807	0.6391	0.6952	0.3598	0.9040	0.7359
KAL	**0.04625**	**0.04875**	**0.04376**	**0.05098**	**0.05531**	**0.06514**	**0.05696**	**0.06571**	**0.04784**	**0.06764**	-0.00098	0.00263	-0.00015		0.6112	0.9350	0.4479	0.8022	0.8546
ROK	**0.04421**	**0.04208**	**0.04765**	**0.04723**	**0.04821**	**0.05775**	**0.05212**	**0.05761**	**0.04496**	**0.06467**	0.00125	0.00569	0.00025	0.00014		0.4077	0.1774	0.5060	0.5122
SIL	**0.05210**	**0.05592**	**0.04834**	**0.05712**	**0.06196**	**0.07315**	**0.06131**	**0.07157**	**0.05216**	**0.07275**	0.00278	0.00391	-0.00013	-0.00267	0.00158		0.8118	0.9632	0.9966
SIN	**0.05825**	**0.06367**	**0.05326**	**0.06152**	**0.06260**	**0.08075**	**0.06561**	**0.07556**	**0.05906**	**0.08079**	0.00275	0.00095	0.00272	0.00169	0.00425	-0.00116		0.3621	0.7691
BAT	**0.04646**	**0.04902**	**0.04760**	**0.05093**	**0.05539**	**0.06434**	**0.05638**	**0.06561**	**0.05160**	**0.06558**	0.00205	0.00191	-0.00173	-0.00080	0.00134	-0.00277	0.00311		0.7486
RUS	**0.04962**	**0.05224**	**0.05398**	**0.05856**	**0.06229**	**0.07233**	**0.06432**	**0.07053**	**0.05454**	**0.07173**	0.00360	0.00343	-0.00045	-0.00206	0.00088	-0.00577	-0.00096	-0.00034	

Table 3. Pairwise genetic distances (F_{ST}) between the 8 Adriatic-Ionian population samples based on Adriatic-Ionina dataset (811 polymorphic loci). F_{ST} indices and p-values are reported below and above the diagonal, respectively; significant indices in bold. Benjamini & Hochberg correction for multiple tests was applied. See Table 1 for location acronyms.

	OTH-A	KAP	BOK	KOR	TRE	TOG	OTR	POC
OTH-A		0.3084	**0.0023**	0.2309	0.1383	0.4631	0.0497	0.0364
KAP	0.00419		**0.0009**	0.0296	0.5433	0.4056	0.1269	0.0375
BOK	**0.00795**	**0.01133**		0.1677	**<0.0001**	**0.0140**	**<0.0001**	**<0.0001**
KOR	0.00407	0.00878	0.00464		**0.0001**	0.7508	**0.0105**	0.0264
TRE	0.00440	0.00265	**0.01154**	**0.01097**		0.7715	**0.0005**	0.0181
TOG	0.00291	0.00422	**0.00787**	0.00138	0.00097		0.2105	0.3810
OTR	0.00609	0.00674	**0.01252**	**0.00843**	**0.00960**	0.00523		0.0863
POC	0.00579	0.00740	**0.01443**	0.00669	0.00655	0.00329	0.00535	

Table 4. Pairwise genetic distances (F_{ST}) between the 9 population samples from the Black Sea based on the Black Sea's dataset (998 polymorphic loci). F_{ST} indices and p-values are reported below and above the diagonal, respectively. Benjamini & Hochberg correction for multiple tests was applied. See Table 1 for location acronyms.

	KAR	TAR	COS	KAL	ROK	SIL	SIN	BAT	RUS
KAR		0.5039	0.5200	0.9232	0.5624	0.8246	0.3893	0.2206	0.1632
TAR	0.00305		0.2894	0.5105	0.5045	0.8897	0.1232	0.2061	0.3549
COS	0.00247	0.0045		0.5670	0.7575	0.4215	0.0930	0.7892	0.4902
KAL	0.00017	0.00318	0.0023		0.7817	0.7631	0.1285	0.7395	0.5749
ROK	0.00217	0.00318	0.00143	0.00118		0.9188	0.2941	0.1630	0.5437
SIL	0.00104	0.00087	0.00302	0.00147	0.00032		0.7063	0.8643	0.7254
SIN	0.00369	0.00698	0.00609	0.00555	0.00426	0.00225		0.4197	0.6066
BAT	0.00405	0.00536	0.00140	0.00158	0.00458	0.00096	0.00386		0.2842
RUS	0.00475	0.00441	0.00302	0.00243	0.00254	0.00185	0.00281	0.00411	

Figure Legends

Figure 1. Map of sampling sites in Mediterranean and Black seas. See table 1 for acronyms.

Figure 2. Discriminant Analysis of Principal Components (DAPC) for the overall dataset of 19 population samples from Mediterranean and Black seas. A) BIC value. Graph of BIC values for increasing value of number of clusters (k). **B)** Graphical table showing the number of original populations (ori) VS the number of inferred clusters (inf) and the group sizes of each correspondence. Original populations legend: ori 1: Corfù Island, Greece (OTH-A); ori 2: Karaburun Peninsula, Albania (KAP); ori 3: Boka Kotorska Bay, Montenegro (BOK); ori 4: Kornati islands, Croatia (KOR); ori 5: Tremiti islands, Italy (TRE); ori 6: Torre Guaceto, Italy (TOG); ori 7: Otranto, Italy (OTR); ori 8: Porto Cesareo, Italy (POC); ori 9: Tunisia (TUN); ori 10: France (FRN); ori 11: Karadag MAP, Ukraine (KAR); ori 12: Tharankut, Ukraine (TAR); ori 13: Costinesti Monastery, Romania (COS); ori 14: Cape Kaliakra, Bulgaria (KAL); ori 15: Ropotamo-Kiten, Bulagiar (ROK); ori 16: Sile, Turkey (SIL); ori 17: Sinop, Turkey (SIN); ori 18: Batumi, Georgia (BAT); ori 19: Russia (RUS). **C)** 2D scatterplot representing the density the inferred numbers of clusters (k=2) on the single discriminant function. **D)** 2D scatterplot representing density of the 19 original populations on the single discriminant function.

Figure 3. Bayesan clustering performed with STRUCTURE on the overall dataset of 19 population samples from Mediterranean and Black seas. Legend: 1: Corfù Island, Greece (OTH-A); 2: Karaburun Peninsula, Albania (KAP); 3: Boka Kotorska Bay, Montenegro (BOK); 4: Kornati islands, Croatia (KOR); 5: Tremiti islands, Italy (TRE); 6: Torre Guaceto, Italy (TOG); 7: Otranto, Italy (OTR); 8: Porto Cesareo, Italy (POC); 9: Tunisia (TUN); 10: France (FRN); 11: Karadag MAP, Ukraine (KAR); 12: Tharankut, Ukraine (TAR); 13: Costinesti Monastery, Romania (COS); 14: Cape Kaliakra, Bulgaria (KAL); 15: Ropotamo-Kiten, Bulgaria (ROK); 16: Sile, Turkey (SIL); 17: Sinop, Turkey (SIN); 18: Batumi, Georgia (BAT); 19: Russia (RUS).

Figure 4. Discriminant Analysis of the Principal Components (DAPC) for the Black Sea's dataset of 9 population samples. A) Graph of BIC values for increasing value of number of clusters (k). **B)** Scatterplot of the discriminant analysis set on the original populations. **C)** 2D scatterplot showing the density of the original populations plotted on the single discriminant. For the acronym locations see Table 1.

Figure 5. Discriminant Analysis of the Principal Components (DAPC) for the Adriatic-Ionian dataset of 8 population samples. A) Graph of BIC values for increasing value of number of clusters (k). **B)** Scatterplot of the discriminant analysis set on the original populations. **C)** 2D scatterplot showing the density of the original populations plotted on the single discriminant. For the acronym locations see Table 1.

Figures

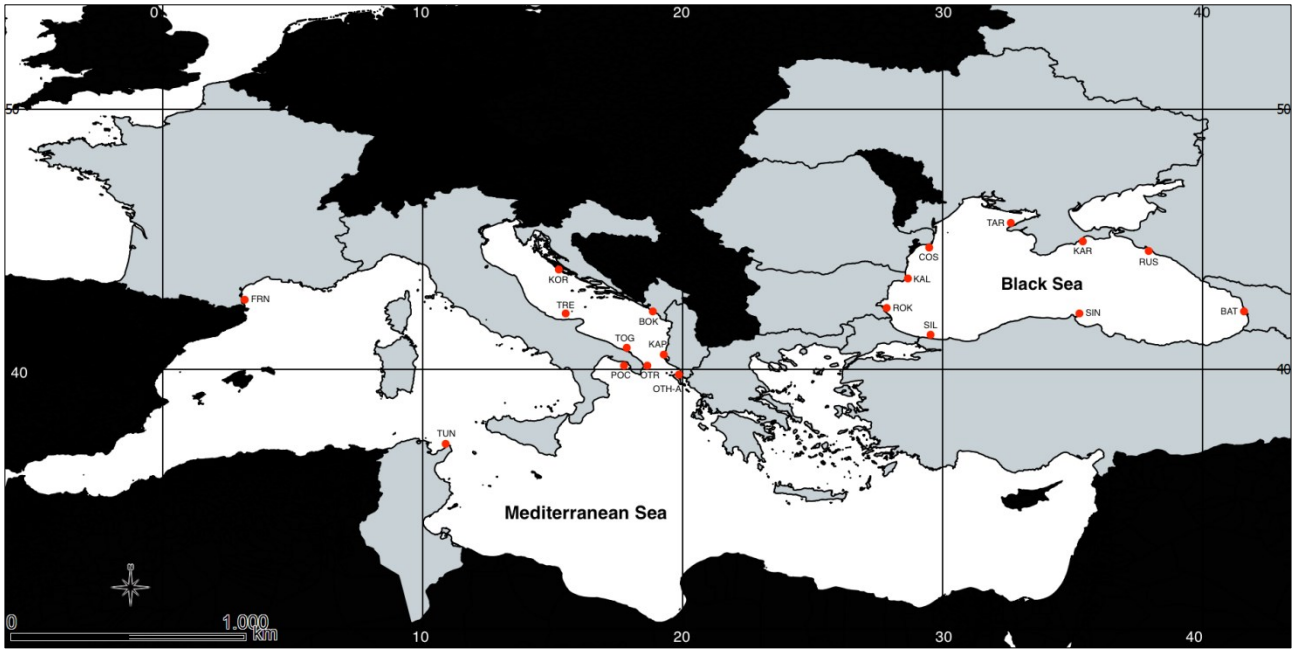


Figure 1. Map of sampling sites in Mediterranean and Black seas. See table 1 for acronyms.

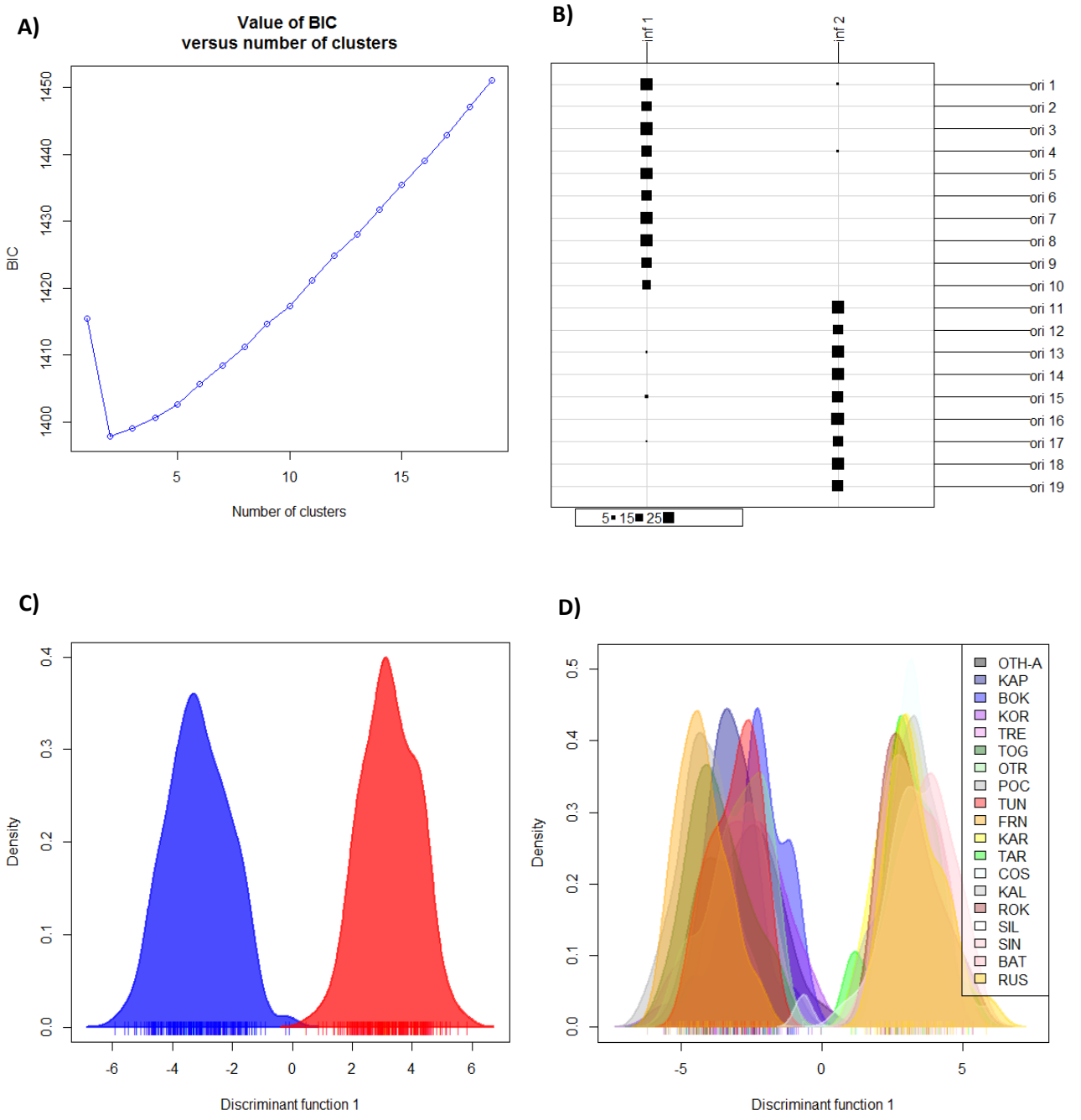


Figure 2. Discriminant Analysis of Principal Components (DAPC) for the overall dataset of 19 population samples from Mediterranean and Black seas. A) BIC value. Graph of BIC values for increasing value of number of clusters (k). **B)** Graphical table showing the number of original populations (ori) VS the number of inferred clusters (inf) and the group sizes of each correspondence. Original populations legend: ori 1: Corfù Island, Greece (OTH-A); ori 2: Karaburun Peninsula, Albania (KAP); ori 3: Boka Kotorska Bay, Montenegro (BOK); ori 4: Kornati islands, Croatia (KOR); ori 5: Tremiti islands, Italy (TRE); ori 6: Torre Guaceto, Italy (TOG); ori 7: Otranto, Italy (OTR); ori 8: Porto Cesareo, Italy (POC); ori 9: Tunisia (TUN); ori 10: France (FRN); ori 11: Karadag MAP, Ukraine (KAR); ori 12: Tharankut, Ukraine (TAR); ori 13: Costinesti Monastery, Romania (COS); ori 14: Cape Kaliakra, Bulgaria (KAL); ori 15: Ropotamo-Kiten, Bulagiar (ROK); ori 16: Sile, Turkey (SIL); ori 17: Sinop, Turkey (SIN); ori 18: Batumi, Georgia (BAT); ori 19: Russia (RUS). **C)** 2D scatterplot representing the density the inferred numbers of clusters (k=2) on the single discriminant function. **D)** 2D scatterplot representing density of the 19 original populations on the single discriminant function.

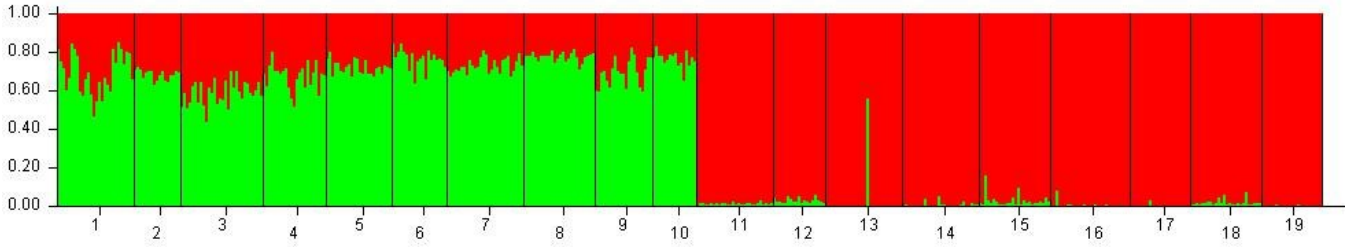


Figure 3. Bayesan clustering performed with STRUCTURE on the overall dataset of 19 population samples from Mediterranean and Black seas. Legend: 1: Corfù Island, Greece (OTH-A); 2: Karaburun Peninsula, Albania (KAP); 3: Boka Kotorska Bay, Montenegro (BOK); 4: Kornati islands, Croatia (KOR); 5: Tremiti islands, Italy (TRE); 6: Torre Guaceto, Italy (TOG); 7: Otranto, Italy (OTR); 8: Porto Cesareo, Italy (POC); 9: Tunisia (TUN); 10: France (FRN); 11: Karadag MAP, Ukraine (KAR); 12: Tharankut, Ukraine (TAR); 13: Costinesti Monastery, Romania (COS); 14: Cape Kaliakra, Bulgaria (KAL); 15: Ropotamo-Kiten, Bulgaria (ROK); 16: Sile, Turkey (SIL); 17: Sinop, Turkey (SIN); 18: Batumi, Georgia (BAT); 19: Russia (RUS).

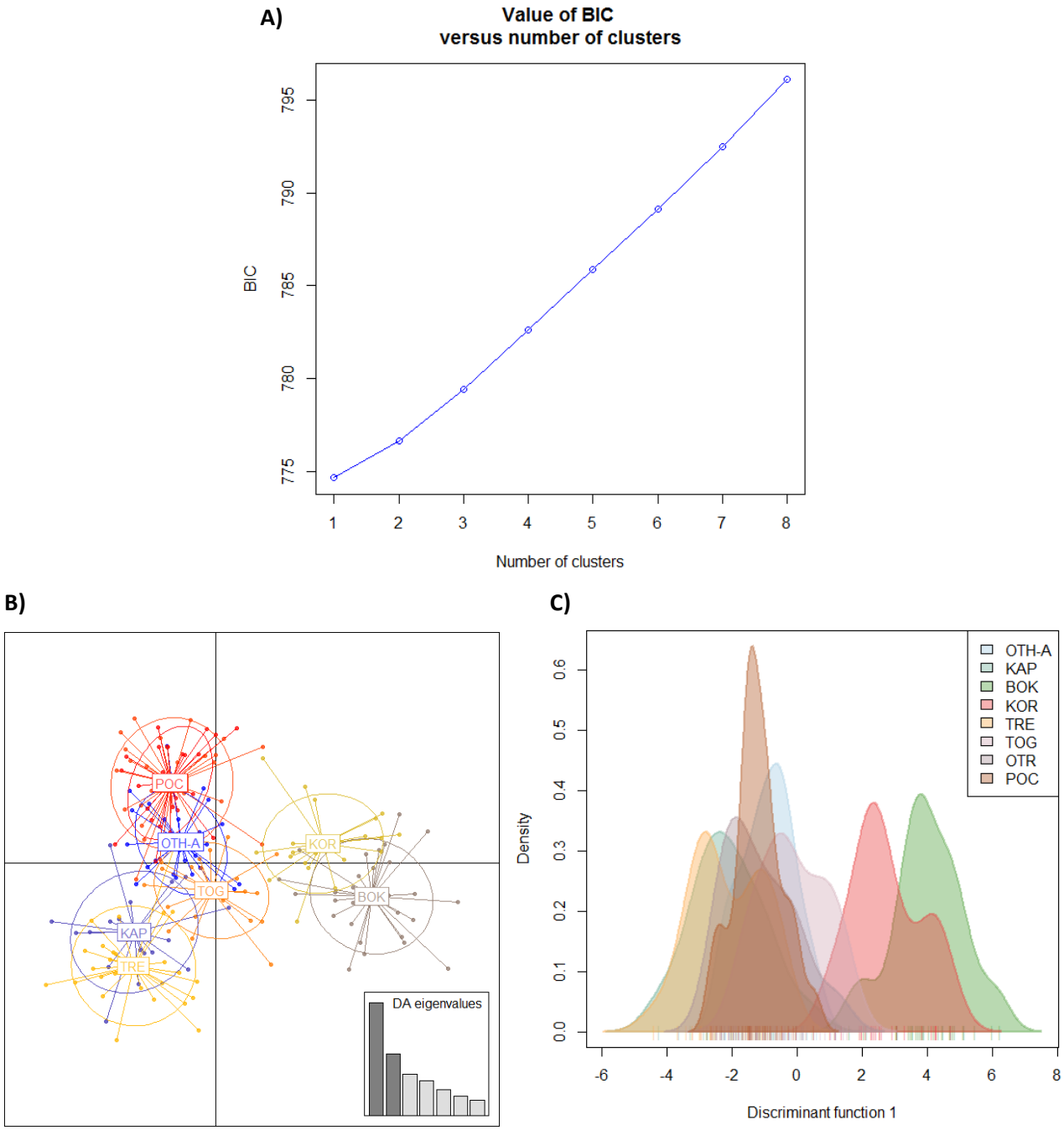


Figure 4. Discriminant Analysis of the Principal Components (DAPC) for the Adriatic-Ionian dataset of 8 population samples. A) Graph of BIC values for increasing value of number of clusters (k). **B)** Scatterplot of the discriminant analysis set on the original populations. **C)** 2D scatterplot showing the density of the original populations plotted on the single discriminant. For the acronym locations see Table 1.

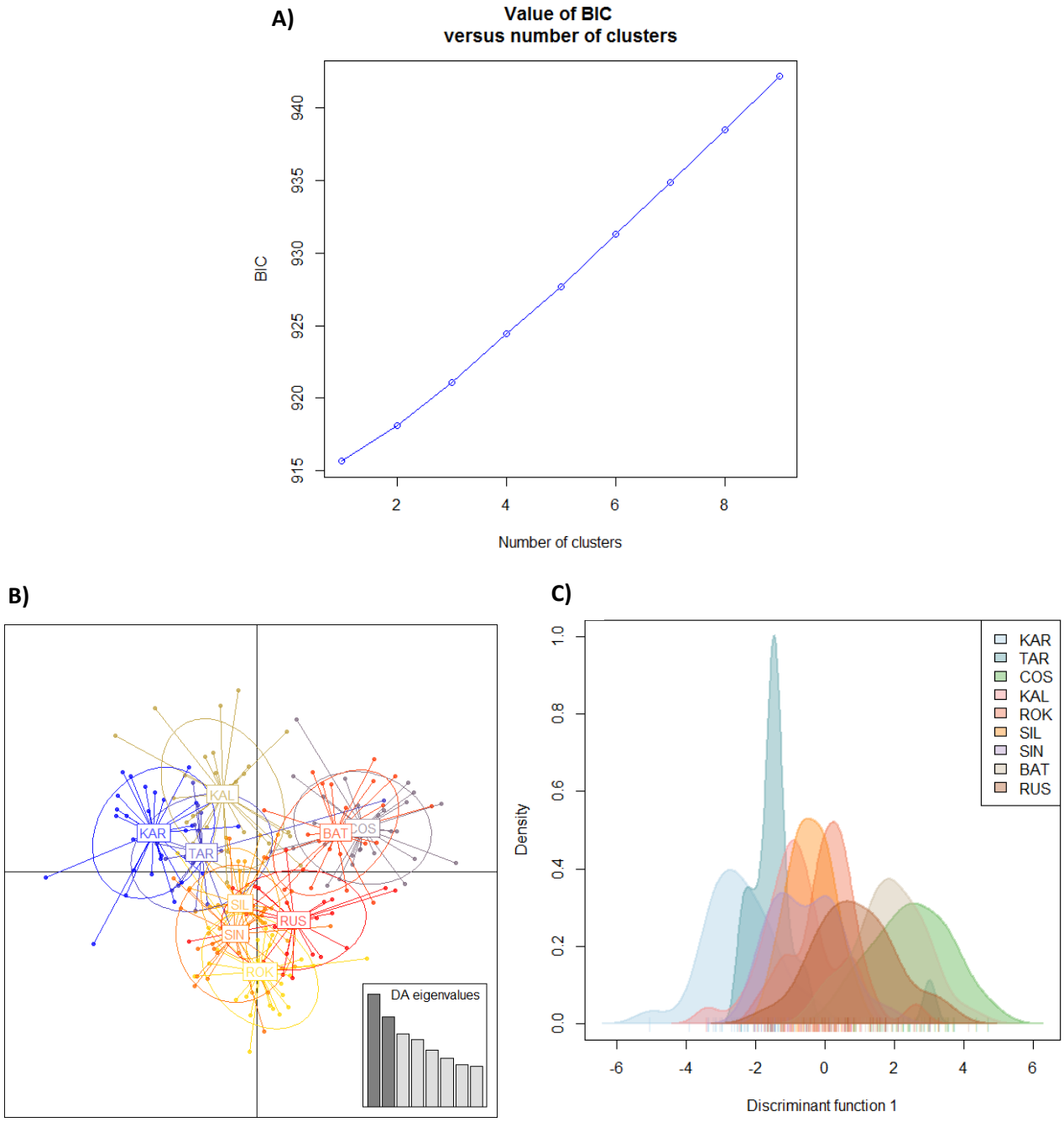


Figure 5. Discriminant Analysis of the Principal Components (DAPC) for the Black Sea's dataset of 9 population samples. A) Graph of BIC values for increasing value of number of clusters (k). **B)** Scatterplot of the discriminant analysis set on the original populations. **C)** 2D scatterplot showing the density of the original populations plotted on the single discriminant. For the acronym locations see Table 1.

Supplementary Material

Table S2. Genetic variability at 512 polymorphic loci (overall dataset) of 19 *Mytilus galloprovincialis* population samples from Mediterranean and Black seas. For each population samples, acronym, sample size (N), obsereved heterozigosity (H_O), unbiased expected heterozigosity (H_E) and allelich richness (A_R) are reported. The allelic richness is calculated based on the minimum population sample size (16).

Acronym	N	H_O	H_E	A_R
OTH-A	28	0.0763	0.0831	1.30
KAP	17	0.0719	0.0778	1.29
BOK	30	0.0723	0.0778	1.28
KOR	23	0.0734	0.0793	1.29
TRE	24	0.0736	0.0810	1.30
TOG	20	0.0682	0.0750	1.27
OTR	28	0.0702	0.0788	1.29
POC	26	0.0762	0.0810	1.30
TUN	21	0.0709	0.0821	1.30
FRN	16	0.0742	0.0842	1.31
KAR	28	0.0713	0.0760	1.28
TAR	19	0.0720	0.0752	1.27
COS	28	0.0703	0.0752	1.28
KAL	28	0.0739	0.0771	1.27
ROK	26	0.0733	0.0768	1.28
SIL	29	0.0697	0.0743	1.26
SIN	22	0.0659	0.0717	1.26
BAT	26	0.0705	0.0774	1.28
RUS	22	0.0710	0.0744	1.27

Table S3. Genetic variability at 811 polymorphic loci (Adriatic-Ionian dataset) of 8 *Mytilus galloprovincialis* populations samples from Adriatic and Ionian seas. For each population samples, acronym, sample size (N), obsereved heterozigosity (H_O), unbiased expected heterozigosity (H_E) and allelich richness (A_R) are reported. The allelic richness is calculated based on the minimum population sample size (18).

Acronym	N	H_E	H_O	A_R
OTH-A	28	0.1095	0.0973	1.41
KAP	18	0.1045	0.0927	1.40
BOK	30	0.1085	0.0950	1.40
KOR	23	0.1086	0.0962	1.41
TRE	26	0.1087	0.0982	1.41
TOG	21	0.1065	0.0942	1.40
OTR	28	0.1067	0.0908	1.40
POC	27	0.1100	0.0993	1.43

Table S4. Genetic variability at 998 polymorphic loci (Black Sea's dataset) of 9 *Mytilus gallorovincialis* populations samples from Black Sea. For each population samples, acronym, sample size (N), obsereved heterozigosity (H_O), unbiased expected heterozigosity (H_E) and allelich richness (A_R) are reported. The allelic richness is calculated based on the minimum population sample size (19).

Acronym	N	H_E	H_O	A_R
KAR	28	0.1059	0.0934	1.41
TAR	19	0.1017	0.0891	1.40
COS	28	0.1065	0.0927	1.40
KAL	28	0.1054	0.0926	1.40
ROK	26	0.1064	0.0943	1.41
SIL	29	0.1058	0.0909	1.41
SIN	22	0.1020	0.0875	1.40
BAT	26	0.1036	0.0899	1.40
RUS	22	0.1058	0.0940	1.41

CHAPTER 4

Seascape genetics of the marbled crab *Pachygrapsus marmoratus* in the Central Mediterranean

Under revision to the journal *Molecular Ecology* as:

Marino IAM, Schiavina M, Aglieri G, Bevilacqua S, Boscari E, Congiu L, Faggion S, Kruschel C, Papetti C, Patarnello T, Paterno M, Voutsinas E, Zane L, Melià P. **Combining genetic data and Lagrangian simulations to assess connectivity patterns of the marbled crab *Pachygrapsus marmoratus* in the Adriatic and Ionian seas.**

Combining genetic data and Lagrangian simulations to assess connectivity patterns of the marbled crab *Pachygrapsus marmoratus* in the Adriatic and Ionian seas

Marino IAM[1,2,+], Schiavina M[2,3,+], Aglieri G[2,4], Bevilacqua S[2,4], Boscari E[1,2], Congiu L[1,2], Faggion S[1,5], Kruschel C[6], Papetti C[1,2], Patarnello T[2,5], Paterno M[1,2], Voutsinas E[7], Zane L[1,2*], Melià P[2,3]

[1]Department of Biology, University of Padova, via U. Bassi 58/B, I-35131 Padova, Italy

[2]Consorzio Nazionale Interuniversitario per le Scienze del Mare, Piazzale Flaminio 9, I-00196 Roma, Italy

[3]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, via Ponzio 34/5, I-20133 Milano, Italy

[4]Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, Università del Salento, Strada Provinciale Monteroni, I-73100 Lecce, Italy

[5]Dipartimento di Biomedicina Comparata e Alimentazione, Università di Padova, viale dell'Università 16, 35020 Agripolis, Legnaro, Padova, Italy

[6]University of Zadar, Ul. Mihovila Pavlinovića, 23000 Zadar, Croatia

[7]Hellenic Centre for Marine Research, 46.7 km Athens Souniou Ave., PO box 712, 19013 Anavyssos, Greece

[+]Equally contributed to this work.

[*]Corresponding author: Lorenzo Zane
Department of Biology, University of Padova, via U. Bassi 58/B, I-35131 Padova, Italy,
Phone: +390498276220, Fax: +390498276209, E-mail: lorenzo.zane@unipd.it

Keywords: Lagrangian simulations, marbled crab, microsatellites, outliers, seascape genetics

Running title: Connectivity patterns of the marbled crab

Abstract

Seascape connectivity plays a crucial role in determining marine population dynamics and genetic structuring. Here we used a multidisciplinary approach to investigate the connectivity of the marbled crab (*Pachygrapsus marmoratus*) in the Adriatic and Ionian basins. We combined genetic analyses (based on a panel of 15 microsatellite loci screened on 314 specimens) and Lagrangian simulations (obtained with a biophysical model of larval dispersal) to explore how currents shape the genetic structure by influencing larval dispersal among eight different locations including existing or planned Marine Protected Areas. We detected an overall genetic homogeneity of marbled crab populations in the study area, largely congruent with the outputs of the Lagrangian simulations, thus pointing out a high level of connectivity among sites. Genetic homogeneity was evident only after removing a single outlier locus, potentially subject to directional selection, that indicated a much higher genetic divergence than the remaining 14 loci. For this reason, we recommend that neutrality tests are routinely performed when investigating populations with weak genetic differentiation. Our results indicate that rocky shore habitats are highly connected within the study area, at least for wide-dispersing species such as *P. marmoratus*.

Extended Summary

Pachygrapsus marmoratus (Brachyura: Grapsidae, Fabricius, 1787) is a semi-terrestrial crab occurring in the intertidal zone of rocky shores. Its distribution range includes the entire Mediterranean basin, the Black Sea and the eastern Atlantic coasts of Europe and Africa, to the Canary and Azores islands (Zariquiey Alvarez 1968; Ingle 1980). The adult stage is relatively sedentary (Cannicci et al. 1999) and the planktonic larval stage lasts about four weeks in the column before the settlement, that occurs between June and October (Drake et al. 1998; Cuesta & Rodriguez 2000; Flores et al. 2002) and seems to follow a semilunar cycle in correspondence with spring tides. Previous genetic studies based on microsatellites revealed weak significant genetic heterogeneity among populations separated by a few hundred km in other Atlantic and Mediterranean regions (Fratini et al. 2008; Silva et al. 2009; Fratini et al. 2011, 2013; Deli et al. in press). This study aimed to understand how ocean currents shape the genetic structure of the marbled crab *P. marmoratus* by mediating larval dispersal among eight different locations in the Adriatic-Ionian basins using a multidisciplinary approach that combines genetic analysis and Lagrangian simulations of dispersal.

The genetic analyses, based on the total 15 microsatellite loci genotyped in 314 samples, revealed high molecular variability in the Adriatic-Ionian seas. But interestingly, removing the locus pm108, identified as a putative outlier by 2 neutrality tests, a general homogeneity among the Adriatic-Ionian sampling sites was detected by the remaining 14 putatively neutral loci. Indeed, the total genetic variation was found within individuals and all fixation indices of the pairwise comparisons between population samples were no more

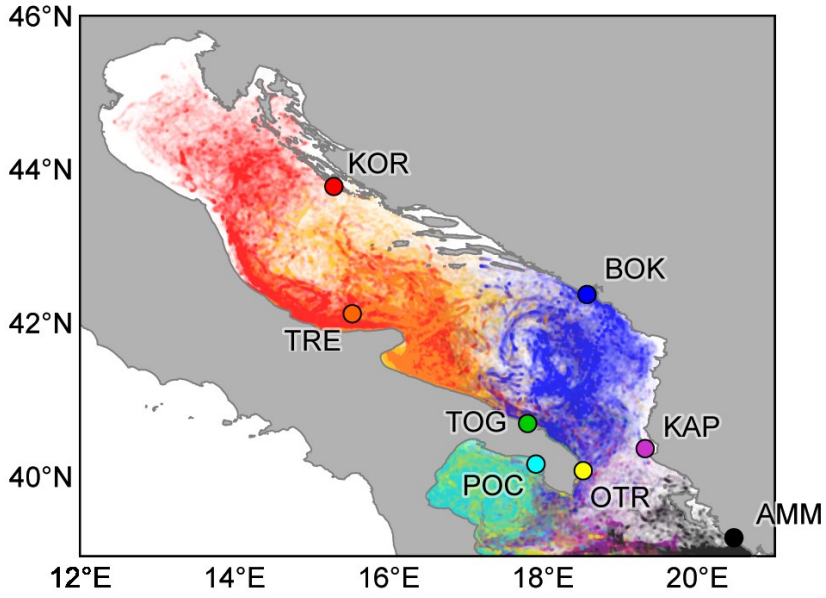
significant (after Benjamini & Hochberg correction) ranging from negative values to a maximum of 0.0111 (Table 1).

Table 1. *Pachygrapsus marmoratus* pairwise genetic differentiation based on 14 neutral loci. Estimates of pairwise F_{ST} among eight population samples are reported below the diagonal; associated P-values are provided above the diagonal. Significant values at the nominal threshold of α=0.05 and corresponding F_{ST} are reported in bold. No values were significant after correction for multiple tests (Benjamini & Hochberg 1995).

	AMM	KAP	BOK	KOR	TRE	TOG	OTR	POC
AMM	–	0.0645	0.0543	0.3307	**0.0470**	**0.0168**	0.0977	0.1782
KAP	0.0085	–	0.9167	0.4399	0.3276	0.3527	0.2720	**0.0280**
BOK	0.0062	-0.0047	–	0.1766	0.1179	0.1659	0.0625	0.0791
KOR	0.0016	0.0015	0.0048	–	0.1850	0.2028	0.1552	**0.0187**
TRE	**0.0059**	0.0031	0.0058	0.0038	–	0.9366	0.4645	**0.0385**
TOG	**0.0066**	0.0041	0.0048	0.0039	-0.0026	–	0.9203	0.1530
OTR	0.0043	0.0051	0.0071	0.0046	0.0013	-0.0018	–	0.0685
POC	0.0026	**0.0111**	0.0056	**0.0078**	**0.0061**	0.0033	0.0050	–

Potential connectivity of *P. marmoratus* among the locations of genetic sampling was investigated via Lagrangian simulations. To this end, we used a biophysical model similar to that developed by Schiavina et al. (2014) to describe the dispersion of the Mediterranean shore crab *Carcinus aestuarii* in the Adriatic and Ionian basins. Lagrangian particles, representing individual zoeae (larvae), were released from the eight sampling locations; simulations were run over the period 2006–2013 to cover the temporal range in which sampled individuals were likely born (assuming an age at capture between 0 and 8 years). Release time was set so as to fit the available information regarding the duration of the zoea stage (1 month, Drake et al. 1998) and the subsequent settlement of *P. marmoratus megalopae* (which peaks between June and October; Cuesta & Rodriguez 2000; Flores et al. 2002). At each spawning event, Lagrangian particles (625 000 per site per year) were released according to a 2-D Gaussian distribution within a 1-km radius from each site, with a uniform distribution between 0.5 and 1 m depth along the vertical direction (to mimic the neustonic habit of *P. marmoratus* larvae, Flores et al. 2002; dos Santos et al. 2008). Figure 2 summarizes the results of the biophysical simulations and shows the trajectories followed by Lagrangian particles across the study area over the simulation period (2006–2013). In summary, patterns of Lagrangian connectivity reflect the counter-clockwise (cyclonic) flow of the oceanic circulation in the study area. Dispersal was predominantly oriented from South to north along the eastern coast of the basin, and southwards along the western coast and in addition, strong westward connections are evident, particularly within the northern South Adriatic (for instance BOK to TOG), and west to east connections are present, though much weaker, within the southern South Adriatic (see for instance OTR to KAP).

Figure 1. Larval dispersal resulting from Lagrangian simulations. Coloured circles indicate the locations of larval release. Legend: black: AMM (Ammoudia, Greece); violet: KAP (Karaburun Peninsula, Albania); blue: BOK (Boka Kotorska Bay, Montenegro); red: KOR (Kornati islands, Croatia); orange: TRE (Tremiti islands, Italy); green: TOG (Torre Guaceto, Italy); yellow: OTR (Otranto, Italy); cyan: POC (Portco Cesareo, Italy). Each dot represents the arrival location of a particle released from each site over the time horizon of the simulation (2006–2013).



To concluding, the results of this study suggest the existence of a fairly homogeneous (from the genetic viewpoint) population of marbled crab in the study area, largely congruent with the outputs of the Lagrangian simulations, which point out a high level of exchange among sites. The genetic data was explained with the existence of random fluctuations of allele frequencies (depending on fecundity, mortality or reproductive success) leading to a scenario of genetic patchiness in a general context of genetic homogeneity, yet previously reported for this species in other studies (Silva et al. 2009; Fratini et al. 2011, 2013). This result is fully in line with the outputs of Lagrangian simulations, pointing out a high level of propagules exchange among sites, although the presence of seasonal and interannual variation in the oceanographic connectivity intensity. Interestingly, we identified a single outlier locus potentially under selection, whose inclusion in the analysis would produce a different picture, largely inflating genetic differentiation estimates. Thus, our results indicate a high level of connectivity between existing and planned MPAs considered in this study (at least for this species, characterized by a relatively long larval duration), and strongly prompt for the use of outlier tests in studies dealing with weak genetic differentiation.

References

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, 289–300.

Cannicci S, Paula J, Vannini M (1999) Activity pattern and spatial strategy in *Pachygrapsus marmoratus* (Decapoda: Grapsidae) from Mediterranean and Atlantic shores. *Marine Biology*, 133, 429–435.

Cuesta JA, Rodriguez A (2000) Zoeal stages of the intertidal crab *Pachygrapsus marmoratus* (Fabricius, 1787) (Brachyura, Grapsidae) reared in the laboratory. *Hydrobiologia*, 436, 119–130.

Deli T, Fratini S, Ragionieri L, Said K, Chatti N, Schubart CD (2016) Phylogeography of the marbled crab *Pachygrapsus marmoratus* (Decapoda, Grapsidae) along part of the African Mediterranean coast reveals genetic homogeneity across the Siculo-Tunisian Strait versus heterogeneity across the Gibraltar Strait. *Marine Biology Research*, in press.

dos Santos A, Santos AMP, Conway DVP, Bartilotti C, Lourenço P, Queiroga H (2008) Diel vertical migration of decapod larvae in the Portuguese coastal upwelling ecosystem: implications for offshore transport. *Marine Ecology Progress Series*, 359, 171–183.

Drake P, Arias AM, Rodriguez A (1998) Seasonal and tidal abundance patterns of decapod crustacean larvae in a shallow inlet (SW Spain). *Journal of Plankton Research*, 20, 585–601.

Flores AAV, Cruz J, Paula J (2002) Temporal and spatial patterns of settlement of brachyuran crab megalopae at a rocky coast in Central Portugal. *Marine Ecology Progress Series*, 229, 207–220.

Fratini S, Ragionieri L, Cutuli G, Vannini M, Cannicci S (2013) Pattern of genetic isolation in the crab *Pachygrapsus marmoratus* within the Tuscan Archipelago (Mediterranean Sea). *Marine Ecology Progress Series*, 478, 173–183.

Fratini S, Schubart CD, Ragionieri L (2011) Population genetics in the rocky shore crab *Pachygrapsus marmoratus* from the western Mediterranean and eastern Atlantic: complementary results from mtDNA and microsatellites at different geographic scales. In: Held C, Koenemann S, Schubart CD (Eds) *Phylogeography and Population Genetic in Crustacea*, Taylor & Francis/CRC Press, Boca Raton, pp. 191–213.

Fratini S, Zane L, Ragionieri L, Vannini M, Cannicci S (2008) Relationship between heavy metal accumulation and genetic variability decrease in the intertidal crab *Pachygrapsus marmoratus* (Decapoda; Grapsidae). *Estuarine, Coastal and Shelf Science*, 79, 679–686.

Ingle RW (1980) British crabs (BM Natural History, Ed). *Oxford University Press*, London.

Schiavina M, Marino IAM, Zane L, Melià P (2014) Matching oceanography and genetics at the basin scale. Seascape connectivity of the Mediterranean shore crab in the Adriatic Sea. *Molecular Ecology*, 23, 5496–5507.

Silva IC, Mesquita N, Schubart CD, Judite Alves M, Paula J (2009) Genetic patchiness of the shore crab *Pachygrapsus marmoratus* along the Portuguese coast. *Journal of Experimental Marine Biology and Ecology*, 378, 50–57.

Zariquiey Alvarez R (1968) Crustaceos decapodos Ibericos. *Investigacion Pesquera*, 32, 1–510.

CHAPTER 5

Discussion and conclusion

Discussion and conclusion

Overall, connectivity data obtained so far in the CoCoNET framework revealed a heterogeneous mosaic of species-specific connectivity patterns throughout the targeted areas in Mediterranean and Black seas.

In this thesis, we focused on 3 long-dispersing species, the sea urchin *Paracentrotus lividus*, the Mediterranean mussel *Mytilus galloprovincialis* and the marble crab *Pachygrapsus marmoratus*. All of them have larval dispersal with a long PLD (up to 4 weeks) and adults sessile (mussel) or with a very limited mobility (sea urchin and crab). The connectivity patterns between the chosen localities in the Central Mediterranean Sea, even though similar, revealed peculiar species-specific features.

The analysis of *P. lividus* population samples provided a clear support for genetic homogeneity within and between Adriatic-Ionian basins. This outcome is broadly consistent with Lagrangian simulations, that predicted a potential larval exchange among the Adriatic-Ionian localities through a relatively persistent flux of propagules from each sites to its closest neighbor, mainly in a counterclockwise direction (Paterno et al. 2017). Similarly, population samples of *P. marmoratus* collected from the same area resulted to be fairly genetic homogenous. After removing a locus putatively outlier under directional selection, a weak signal of differentiation was detected, probably indicative of genetic patchiness. This data was explained with random fluctuations of allele frequencies (depending on fecundity, mortality or reproductive success) leading to a scenario of genetic patchiness, also previously reported (Silva et al. 2009; Fratini et al. 2011, 2013), in a context of genetic homogeneity. This result is fully in line with the outputs of Lagrangian simulations, pointing out a high level of propagules exchange among sites, although the presence of seasonal and interannual variation in the oceanographic connectivity intensity (Marino et al. under revision). The analysis of the population structure of *M. galloprovincialis* found out a slightly different connectivity picture, indeed genetic heterogeneity was detected in the Adriatic-Ionian basins. Bayesian clustering methods failed in detecting genetically differentiated clusters, likely due to the weakness of the differentiation, but a hierarchical AMOVA found that genetic variation of the dataset could be best partitioned into 3 groups: first group including the Tremiti islands's sample (TRE) alone, the second one comprising Montenegro's sample (BOK) along with the Croatian one (KOR), and at last the 4 Italian samples together with the Greek (OTH-A) and Albanian (KAP) samples. This pattern underlined an East-West differentiation in the Central Adriatic Sea, whereas a general homogeneity was found in the Southern Adriatic and Ionian seas (Paterno et al. in prep).

Moving the attention to other CoCoNET species indicates that the detection of genetic differentiation in the Adriatic-Ionian basins is not uncommon. For the black scorpionfish *Scorpaena porcus*, a benthic species dispersing mainly through the larval stage, an East-West differentiation in the Central Mediterranean seemed to prevail, with a clear genetic break between the eastern and western localities (Boissin et al. 2016). The east Atlantic peacock wrasse *Symphodus tinca* is a demersal fish that displays a territorial behavior, and it's supposed to disperse mostly during its larval phase. Carreras et al (2017) explored the

population structure of this species in the Adriatic and Ionian seas and identified 2 barriers to gene flow, separating the Eastern shore locations from the South-Western one (similarly to what observed in *S. porcus*) and differentiating the Tremiti islands's sample (TRE) from all the others localities. At last, the case of the habitat former *Posidonia oceanica* that showed a marked genetic structure in the same area, even between the closest localities, with a north-to-south subdivision (Jahke et al. in review). The genetic differentiation observed in the Central Mediterranean for these species is likely due to the complexity of the local oceanographic conditions, characterized by the presence of the stable and permanent gyral circulation in the Adriatic Sea, that can act as a barrier in reduction the gene flow between the localities where the population samples of these species were collected.

Taking into account the Black Sea in the CoCoNET framework, only data of connectivity regarding the species *Mytilus galloprovincialis* and *Scorpaena porcus* are so far available. In general, little is known about the genetic structure of the species living here, and the CoCoNET project offered one of the first wide and structured sampling design in this area, covering all the countries bordering the sea. For *M. galloprovincialis*, a totally lack of genetic differentiation was found in the Black Sea, indeed all the population samples resulted to be genetically homogeneous (Paterno et al. in prep). For *S. porcus*, genetic differentiation was detected only for 2 isolated localities without a clear geographic pattern, that the authors attributed to chaotic genetic patchiness (Boissin et al. 2016).

Connectivity data obtained so far highlighted species-specific connectivity patterns, ranging from the total absence of genetic differentiation in the sea urchin *P. lividus*, to the strong genetic structure even at a local scale in the habitat former *P. oceanica*. Despite the ability dispersal of all these species relies on the pelagic larval stage, many species-specific biological traits (e. g. PLD, reproductive timing, the recruitment success) and the interactions with environmental settings and ecological factors, can results in the different connectivity outcomes we observed in these studies. Interestingly, Pascual et al. (2017) reviewed a considerable number of scientific studies to estimate the impact of 2 life history traits affecting the connectivity, the PLD and the adult life strategy, on the population differentiation around seven major oceanographic fronts in the Mediterranean Sea. In general, the authors observed that oceanographic fronts have an effect in reducing the gene flow, but the extent of this effect is not homogeneous and strictly depends on the life traits of the species considered. Specifically, species with adult mobility and a long PLD seems to be mostly affected by the presence of an oceanographic front, indeed a significant connectivity reduction was observed in these species. On the other hand, benthic sessile species, mostly habitat-formers, showed significant genetic differentiation independently of the presence of oceanographic fronts.

Overall, the heterogeneity of connectivity patterns depends on species-specific peculiarities in relation to environmental features, and connectivity results so far obtained warns against any generalization in defining conservation units based on the results of the single species. The current depletion of marine biodiversity and environment strongly calls for effective conservation and management strategies that take

into account the complexity of marine ecosystem. The wide heterogeneity of species-specific connectivity outcomes suggests that the implementation of ecological effective networks of MPAs should standing on meta-community information across a range of spatial scales (Melià et al. 2016) and umbrella species data (Pascual et al. 2017) to obtain a "collection of networks of genetic variation of all species within a community" (Fortuna et al. 2009).

The resilience of marine ecosystem depends on our ability to set effective conservation and management strategies to reducing the alteration of this environment and mitigating the negative effect of anthropic activities. In this fast-degenerating scenario of habitat loss and biodiversity depletion, the restoration and conservation of marine biodiversity are top priorities.

CoCoNET project is providing an invaluable knowledge towards this goals.

References

Boissin E, Micu D, Janczyszyn-Le Goff M, et al. (2016) Contemporary Genetic Structure and Post-Glacial Demographic History of the Black Scorpionfish, *Scorpaena Porcus*, in the Mediterranean and the Black Seas. *Molecular Ecology*, 25, 2195-2209.

Carreras C, Ordóñez V, Zane L, Kruschel C, Nasto I, Macpherson E, Pascual M (2017) Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation. *Scientific Reports*, 7:43417.

Fortuna MA, Albaladejo RG, Fernández L, Aparicio A, Bascompte J (2009) Networks of spatial genetic variation across species. *PNAS*, 106, 19044–19040.

Fratini S, Schubart CD, Ragionieri L (2011) Population genetics in the rocky shore crab *Pachygrapsus marmoratus* from the western Mediterranean and eastern Atlantic: complementary results from mtDNA and microsatellites at different geographic scales. In: Held C, Koenemann S, Schubart CD (Eds) *Phylogeography and Population Genetic in Crustacea*, Taylor & Francis/CRC Press, Boca Raton, pp. 191–213.

Fratini S, Ragionieri L, Cutuli G, Vannini M, Cannicci S (2013) Pattern of genetic isolation in the crab *Pachygrapsus marmoratus* within the Tuscan Archipelago (Mediterranean Sea). *Marine Ecology Progress Series*, 478, 173–183.

Jahnke M, Casagrandi R, Melià P et al. (submitted to *Diversity & Distributions*) Connectivity, Sexual Reproduction and Selection Pressure in the Seagrass P. oceanica and the Implications for Conservation.

Marino IAMM, Schiavina M, Aglieri G et al. (submitted to *Molecular Ecology*) Assessing Connectivity Patterns of the Marbled Crab *Pachygrapsus marmoratus* in the Adriatic and Ionian Seas Combining Genetic Data and Lagrangian Simulations.

Melià P, Schiavina M, Rossetto M, Gatto M, Fraschetti S, Casagrandi R (2016) Looking for Hotspots of Marine Metacommunity Connectivity: a Methodological Framework. *Scientific Reports*, 6, 23705.

Pascual M, Rives B, Schunter C, Macpherson E (2017) Impact of life history traits on gene flow: A multispecies systematic review across oceanographisc barriers in the Mediterranean Sea. *PLos ONE*, 12, e0176419.

Paterno M, Schiavina M, Aglieri G et al. (2017) Population genomics meet Lagrangian simulations: oceanographic patterns and long larval duration ensure connectivity among *Paracentrotus lividus* populations in the Adriatic and Ionian seas. *Ecology & Evolution*, 00, 1–17,

Paterno M, Ben Souissi J, Boscari E, Chassanite A, Congiu L, Guarnieri G, Kruschel C, Macic V, Marino IAM, Micu D, Voutsinas M, Papetti C, Zane L. (in preparation) A genome-wide approach to the global phylogeography of the mussel *Mytilus galloprovincialis* in the Mediterranean and Black Seas.

Silva IC, Mesquita N, Schubart CD, Judite Alves M, Paula J (2009) Genetic patchiness of the shore crab *Pachygrapsus marmoratus* along the Portuguese coast. *Journal of Experimental Marine Biology and Ecology*, 378, 50–57.

Part 2

LANDSCAPE GENETICS IN PARASITIC DISEASES

CHAPTER 6

Background

Background

Neglected Tropical Diseases (NTDs)

Neglected Tropical Diseases (NTDs) are a group of parasitic and bacterial infectious diseases affecting worldwide almost 3 billion people living in poverty, especially in the tropical areas (Hotez et al. 2007). Among the parasitic diseases, it's possible to find out the well-known lymphatic filariasis, schistosomiasis, trypanosomiasis, leishmaniasis and Chagas disease (Who 1990; Michael and Gambhir 2010; Tarleton 2007). The NTDs causative agents have usually a complex life-cycle, involving at least one vertebrate or invertebrate intermediate host, as well as human beings. Although the wide and global distribution, NTDs determine 'patchy' geographic spots where parasite, vector and host conspire to promote an intense local transmission. Disease transmission depends on the interactions and the spatial/temporal heterogeneity of biotic (e.g. vector abundance) and abiotic (e.g. environment) factors, whose predictions could point out the emergence, occurrence, and spread of this group of infectious diseases (Randolph and Rogers 2010; Vazquez-Prokopec et al. 2016).

Landscape genetics in parasitic diseases

A substantial heterogeneity exists in the dispersal ability, distribution and transmission routes of parasitic species. Specifically, dispersal and distribution of NTDs pathogens and their vectors are strongly influenced by ecological settings and landscape variation. Understanding how abiotic and biotic variables interact and affect the movements and distribution of parasites, hosts and vector species is key for predicting and controlling the disease spread (Schwabl et al. 2016). Furthermore, it is essential to estimate the extent to which populations of parasites and their vectors are connected exploring their spatial genetic structuring. The last 10 years have seen an increasing interest in the population genetic structure of pathogens and vectors especially in human diseases (Bourguinat et al. 2007; Koffi et al. 2009; Norton et al. 2010). Given the determinant role of the environment in affecting parasites and vector distributions, it's not surprising that epidemiologists are increasingly taking the emerging field of landscape genetics (Manel et al. 2003) as working framework to elucidate those mechanisms and interactions driving the spatial distribution and spread of parasitic diseases (Archie et al. 2008; Biek and Real 2010).

Chagas disease

The digenic single-cell kinetoplastid *Trypanosoma cruzi* is the most important parasite in Latin America, where it is the infectious agent of Chagas disease. Over nine million people are supposed to be infected each year and Chagas disease caused around 14 000 deaths annually (Rassi et al. 2010). Chagas' vectors are a few dozen triatomines species (Reduviidae, Triatominae) commonly known as kissing bugs (Lent and

Wygodzinsky 1979; Stevens et al. 2011). In Ecuador, where the Chagas disease is endemic in many regions, one of the 2 main insect vectors is *Rhodnius ecuadoriensis*, widely distributed from Coastal and Southern Ecuador to Northern Peru (Grijalva and Villacis 2009; Grijalva et al. 2012). Infection with *T. cruzi* is a zoonosis with numerous wild mammal reservoir hosts, the infection in humans occurs via different transmission routes, mainly through the contact with the infected faeces of domiciliated triatomines (Carlier et al. 2011; Yoshida et al. 2011). Indeed, nevertheless nation-wide insecticide-spraying programs to eradicate vector populations in human dwellings have successfully reduced Chagas' transmission (Dias et al. 2002; Moncayo and Silveira 2009), domicile recolonization occurs thanks to the ability of the insect vectors to disperse from sylvatic to domestic environments and establish local domestic populations (Noireau 2005; Guhl et al. 2009; Ceballos et al. 2011).

The project "Landscape genetics drive Neglected Tropical Diseases interventions: Chagas disease in Ecuador and Peru"

At the University of Glasgow, I was involved in a National Institutes for Health funded project regarding the study of the Chagas disease in Latin America, titled "Landscape genetics drive Neglected Tropical Diseases interventions: Chagas disease in Ecuador and Peru". This project had 3 specific aims to assist with long-term local elimination of Chagas disease transmission in the selected study areas. The 3 aims are briefly reported as follow: 1) the establishment of population genomic structure of the parasite *Tripanozoma cruzi* in Southern Ecuador (Loja Province) and Northern Peru; 2) the establishment of the population genomic structure of the kissing bug *Rhodnius ecuadoriensis*, the main triatomine vector of *T. cruzi* infection to humans in the same area, with a special attention in comparing wild and domestic disease transmission foci; 3) the evaluation of the landscape genetics structure of both *T. cruzi* and *R. ecuadoriensis* for understanding environmental features that influence parasite dispersal and transmission in these areas. The goal was to detect genetic discontinuities within both parasite and vector populations simultaneously, and associate them with landscape and other environmental features in determining spatially associated clines and cryptic boundaries. Loja Province in Ecuador is currently targeted by the Ecuadorian Chagas Disease Control Program (EChP) where, complementing disease prevention efforts, recent progress has been made in understanding local vector dynamics (Grijalva et al. 2011; 2012). Conversely, triatomine control activities are not currently being implemented in Northern Peru. This is due in part to the lack of information regarding vector biology and distribution, and lack of efforts to detect and treat Chagas disease (Vargas et al. 2007). By associating the patterns of genetic discontinuities in parasite and vector with dispersal processes, it will be possible to form the basis for informed regional intervention strategies targeted at Chagas disease control.

In detail, Luis Enrique Hernandez-Castro (PhD student, University of Glasgow) and I worked on a pilot study on the vector *Rhodnius ecuadoriensis*. By using a NGS-based genome-scale approach (2b-RADseq) conducted on 20 specimens collected from four communities of Ecuador, we i) assessed and compared the effectiveness of three type IIB Restriction Enzymes (AlfI, BcgI, CspCI) in cutting *R. ecuadoriensis* genomic DNA samples; ii) rapidly and cost-effectively generated several hundred Single Nucleotide Polymorphism (SNP) markers for *R. ecuadoriensis* with BcgI and CspCI, and iii) performed preliminary genomic analysis to test the efficacy of our panel of SNP markers and obtained a first evaluation of the genetic population structure of this vector in the selected area in Ecuador.

Methods and results of this work were described and discussed in a paper (reported in the following Chapter) recently published on *PLoS Neglected Tropical Diseases* as:

Hernandez-Castro LE, **Paterno M**, Villacís AG, Andersson B, Costales JA, De Noia M, Ocaña-Mayorga S, Yumiseva CA, Grijalva MJ, Llewellyn MS (2017) 2b-RAD Genotyping for Population Genomic Studies of Chagas Disease vectors: *Rhodnius ecuadoriensis* in Southern Ecuador. *PLOS Neglected Tropical Diseases,* 11(7): e0005710. https://doi.org/10.1371/journal.pntd.0005710.

References

Archie EA, Luikart, Ezenwa VO (2008) Infecting epidemiology with genetics: a new frontier in disease ecology. *Trends in Ecology and Evolution*, 24, 21–30.

Biek R, Real LA (2010) The landscape genetics of infectious disease emergence and spread. *Molecular Ecology*, 19, 3515–3531.

Bourguinat C, Pion SD, Kamgno J et al. (2007) Genetic selection of low fertile *Onchocerca volvulus* by ivermectin treatment. *PLoS Neglected Tropical Diseases*, 1, e72.

Carlier Y, Torrico F, Sosa-Estani et al. (2011) Congenital Chagas disease: recommendations for diagnosis, treatment and control of newborns, siblings and pregnant women. *PLoS Neglected Tropical Diseases*, 5, e1250.

Ceballos LA, Piccinali RV, Marcet PL et al. (2011) Hidden sylvatic foci of the main vector of chagas disease *Triatoma infestans*: threats to the vector elimination campaign? *PLoS Neglected Tropical Diseases*, 5, e1365.

Dias J, Silveira A, Schofield C (2002) The impact of Chagas disease control in Latin America: a review. *Mem Inst Oswaldo Cruz*, 97, 603–612.

Grijalva MJ, Villacis AG (2009) Presence of *Rhodnius ecuadoriensis* in sylvatic habitats in the southern highlands (Loja Province) of Ecuador. *Journal of Medical Entomologt*, 46, 708–11.

Grijalva MJ, Villacis AG, Ocana-Mayorga S, Yumiseva CA, Baus EG (2011) Limitations of selective deltamethrin application for triatomine control in central coastal Ecuador. *Parasit Vectors*, 4, 20.

Grijalva MJ, Suarez-Davalos V, Villacis AG, Ocaña-Mayorga S, Dangles O (2012) Ecological factors related to the widespread distribution of sylvatic *Rhodnius ecuadoriensis* populations in southern Ecuador. *Parasit Vectors*, 5, 17.

Guhl F, Pinto N, Aguilera G (2009) Sylvatic triatominae: a new challenge in vector control transmission. *Mem Inst Oswaldo Cruz*, 104, 71–5.

Hotez PJ, Molyneux DH, Fenwick A, Kumaresan J, Sachs SE et al. (2007) Control of neglected tropical diseases. N Engl J Med, 357, 1018–1027.

Lent H, Wygodzinsky PW (1979) Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. Bulletin of the AMNH, 1979, 163, 125–520.

Koffi M, De Meeus T, Bucheton B et al. (2009) Population genetics of *Trypanosoma brucei gambiense*, the agent of sleeping sickness in Western Africa. *PNAS*, 106, 209–214.

Manel S, Schwartz K, Luikart G, Taberlet P (2003) Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, 18, 189–197.

Michael E, Gambhir M (2010) Vector transmission heterogeneity and the population dynamics and control of lymphatic filariasis. *Advances in Experimental Medicine and Biology*, 673, 13–31.

Moncayo Á, Silveira AC (2009) Current epidemiological trends for Chagas disease in Latin America and future challenges in epidemiology, surveillance and health policy. *Mem Inst Oswaldo Cruz*, 104, 17–30.

Noireau F, Cortez M, Monteiro F, Jansen A, Torrico F (2005) Can wild foci in Bolivia jeopardize Chagas disease control efforts? *Trends in Parasitology*, 21, 7–10.

Norton AJ, Gower CM, Lamberton PH et al. (2010) Genetic consequences of mass human chemotherapy for *Schistosoma mansoni*: population structure pre- and post-praziquantel treatment in Tanzania. *The American Society of Tropical Medicine and Hygiene*, 83, 951–957.

Randolph SE, Rogers DJ (2010) The arrival, establishment and spread of exotic diseases: patterns and predictions. *Nature Reviews Microbiology*, 8, 361–371.

Rassi A Jr, Rassi A, Marin-Neto JA (2010) Chagas disease. *Lancet*, 375, 1388-1402.

Schwabl P, Llewellyn MS, Landguth EL et al. (2016) Prediction and Prevention of Parasitic Diseases Using a Landscape Genomics Framework. *Trends in Parasitology*, 33, 264–275.

Stevens L, Dorn PL, Schmidt JO, Klotz JH, Lucero D, Klotz SA (2011) Kissing Bugs. The vectors of Chagas. *Advances in Parasitology*, 169–92.

Tarleton RL, Reithinger R, Urbina JA, Kitron U, Gürtler RE (2007) The challenges of Chagas Disease - grim outlook or glimmer of hope? *PLoS Medicine*, 4, e332.

Vargas F, Cordova Paz Soldan O, Marin C et al. (2007) Epidemiology of American trypanosomiasis in northern Peru. *Annals of Tropical Medicine & Parasitology*, 101, 643–648.

Vazquez-Prokopec GM, Perkins TA, Waller LA et al. (2016) Coupled Heterogeneities and Their Impact on Parasite Transmission and Control. *Trends in Parasitology*, 32, 356–67.

Who (1990) Control of the leishmaniases. *Geneva: World Health Organization*.

Yoshida N, Tyler KM, Llewellyn MS (2011) Invasion mechanisms among emerging food-borne protozoan parasites. *Trends in Parasitology*, 27, 459–466.

CHAPTER 7

Landscape genetics drive Neglected Tropical Diseases interventions: Chagas disease in Ecuador and Peru

Published as:

Hernandez-Castro LE, Paterno M, Villacís AG et al. (2017) **2b-RAD Genotyping for Population Genomic Studies of Chagas Disease vectors: *Rhodnius ecuadoriensis* in Southern Ecuador.** *PLoS Neglected Tropical Diseases*, 11(7): e0005710. https://doi.org/10.1371/journal.pntd.0005710

2b-RAD Genotyping for Population Genomic Studies of Chagas Disease Vectors: *Rhodnius ecuadoriensis* in Southern Ecuador

Luis E Hernandez-Castro [1*¶], Marta Paterno [2,3¶], Anita G Villacís [4], Björn Andersson [5], Jaime A Costales [4], Michele De Noia [6], Sofía Ocaña-Mayorga [4], Cesar A Yumiseva [4], Mario J Grijalva [4,7&], Martin S Llewellyn [1*&]

[1] Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, United Kingdom.

[2] Department of Biology, University of Padua, Padua, Italy.

[3] Consorzio Nazionale Interuniversitario per le Scienze del Mare (CoNISMa), Rome, Italy

[4] Center for Research on Health in Latin America, School of Biological Sciences, Pontifical Catholic University of Ecuador, Quito, Ecuador.

[5] Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden.

[6] Department of Animal Behaviour, Bielefeld University, Bielefeld, Germany.

[7] Infectious and Tropical Disease Institute, Department of Biomedical Sciences, Heritage College of Osteopathic Medicine, Ohio University, Ohio, United States of America.

[*] Corresponding authors

Email: l.hernandez-castro.1@research.gla.ac.uk (LEH-C)

Email: Martin.Llewellyn@glasgow.ac.uk (MSL)

[¶] These authors contributed equally to this work

[&] These authors also contributed equally to this study

Running title: 2b-RAD Genotyping in *Rhodnius ecuadoriensis*

Abstract

Background. *Rhodnius ecuadoriensis* is the main triatomine vector of Chagas disease, American trypanosomiasis, in Southern Ecuador and Northern Peru. Genomic approaches and next generation sequencing technologies have become a powerful tool for investigating population diversity and structure which is a key consideration for vector control. Here we assess the effectiveness of three different 2b-RAD genotyping strategies in *R. ecuadoriensis* to provide sufficient genomic resolution to tease apart microevolutionary processes and undertake some pilot population genomic analyses.

Methodology/Principal Findings. The 2b-RAD protocol was carried out in-house at a non-specialised laboratory using 20 *R. ecuadoriensis* adults collected from the Central coast and Southern Andean region of Ecuador, from June 2006 to July 2013. 2b-RAD sequencing data was performed on an Illumina MiSeq instrument and analysed with the STACKS *de novo* pipeline for loci assembly and Single Nucleotide Polymorphisms (SNPs) discovery. Preliminary population genomic analyses (global AMOVA and Bayesian clustering) were implemented. Our results showed that the 2b-RAD genotyping protocol is effective for *R. ecuadoriensis* and likely for other triatomine species. However only *Bcg*I and *CspC*I restriction enzymes provided a number of markers suitable for population genomic analysis at the read depth we generated. Our preliminary genomic analyses highlighted signal of genetic structuring across the study area.

Conclusions/Significance. Our findings suggest that 2b-RAD genotyping is both a cost effective and methodologically simple approach for generating high resolution genomic data for Chagas disease vectors with the power to distinguish between different vector populations at epidemiologically relevant scales. As such, 2b-RAD represents a powerful tool in the hands of medical entomologists with limited access to specialized molecular biological equipment.

Author Summary

Understanding Chagas disease vector (triatomine bugs) population dispersal is key for the design of control measures tailored for the epidemiological situation of a particular region. In Ecuador, *Rhodnius ecuadoriensis* is a cause of concern for Chagas disease transmission, since it is widely distributed from the Central coast to Southern Ecuador. Here, a wide-genome sequencing (2b-RADseq) approach was performed in 20 specimens from four communities from Manabí (Central coast) and Loja (Southern) provinces in Ecuador, and the effectiveness of three type IIB restriction enzymes was assessed. The findings of this study show that this genotyping methodology is cost effective in *R. ecuadoriensis* and likely in other triatomine species. In addition, preliminary population genomic analysis results detected a signal of population structure among geographically distinct communities and genetic variability within communities. As such, 2b-RADseq shows significant promise as a relatively low-tech solution for determination of vector population genomics, dynamics, and spread.

Introduction

Vector control has been the mainstay of Chagas Disease (CD) control strategies in Latin America. Several Latin American countries implemented nation-wide insecticide-spraying programs to eradicate CD vector populations in human dwellings over the last 30 years. These campaigns resulted in a dramatic reduction in vectorial transmission [1-3]. Despite this success, domicile recolonization is a constant threat due to the ability of several triatomines species to disperse from sylvatic to domestic/peridomestic environments and establish local domestic populations [4-8].

Triatomines, members of the arthropod family Reduviidae, subfamily Triatominae, commonly known as kissing bugs, are distributed from the southern United States to as far away as central Argentina [9]. Over 130 species are identified, but only a few dozen are known to transmit CD [10]. In Ecuador, *Triatoma dimidiata* and *Rhodnius ecuadoriensis* are main vectors of CD, with the latter widely distributed from Coastal and Southern Ecuador to Northern Peru [11,12].

Multiple molecular genetic studies exist which attempt to explain genetic structure and gene flow in triatomine populations [8,13-20]. An example of those tailored to address defined epidemiological hypotheses include that of Fitzpatrick *et al*. [13]. Fitzpatrick *et al*. confirmed that gene flow (and therefore vector dispersal) occurs between sylvatic, domicile and peridomicile ecotopes in Venezuelan *Rhodnius prolixus*, based on pairwise F_{ST} values from both cytochrome b (*cytb*) and nine microsatellites. *R. prolixus* is the major vector species in Venezuela and Colombia, as well as Andean and Central American countries. Fitzpatrick *et al.*'s data suggested that colonization of domestic locales by wild triatomines is indeed possible in the region, and these findings had major implications for control. Other species have also been the subject of study. Population genetic data from *Triatoma infestans* based on ten microsatellites showed fine-scale genetic structure in domestic populations several years after the spraying of insecticides [18]. In this case, genetic data were tested under two different models of dispersal: isolation by distance, and hierarchical island with stratified migration models. The latter best reflected vector genetic structure among the sample sites. Finally, Almeida and colleagues compared *cytb* and 8 microsatellites in *Triatoma brasiliensis* to investigate its genetic structure and to assess gene flow among sylvatic and domestic/peridomestic populations [20]. As with Fitzpatrick *et al.* pairwise comparison of F_{ST} values obtained from microsatellite analysis also demonstrated connectivity between locales [13].

Given that vector control remains the mainstay of CD intervention strategies, greater understanding of vector genetics and dispersal is urgently required. Of particular importance are genotyping approaches that provide very high resolution at local, epidemiologically relevant scales, as well as combinability and shareability of datasets between different studies and research groups. Microsatellite loci offer little flexibility in the in terms of shareability as data standardization guidelines for amplicon size estimation and allele nomenclature between laboratories, although possible [21], are rarely established, time-consuming

and expensive to resolve, an issue already seen in *Trypanosoma cruzi* typing [22]. Likely as a function of funding constraints, molecular genetic research on triatomine vectors and, CD in general, is a very late arrival on the 'omics scene. The belated publication of *R. prolixus* genome last year represents a step in the right direction and has revealed much about the core adaptations that underpin the biological success of triatomines [23]. A number of expressed sequence tags have been developed for *T. infestans* [24,25]. However, in general, genome sequencing efforts in triatomines so far have yielded little benefit to scientists and public health professionals attempting to map vector dispersal.

In tandem with the emergence of high throughput Next Generation Sequencing (NGS) approaches, several groups have pioneered the use of restriction enzymes (REs) on Restriction Site Associated DNA sequencing (RADseq) protocols to allow a small fraction of the genome to be sequenced across multiple samples [26-34]. Several variants of the RADseq technique currently exist [35-39]; however, protocol choice to address a specific research question must balance technical issues, budget and laboratory capacity [40].

The 2b-RAD genotyping strategy specifically uses type IIB restriction enzymes (IIB-REs) for genomic DNA (gDNA) digestion [38]. Advantages of this protocol include simplicity and cost-efficiency, since it is carried out within 3 steps in the same 96-well plate, as compared to 4-6 steps in other RADseq protocols [35-37, 39]. Furthermore, library preparation can be achieved with no more than a PCR machine and a standard agarose gel. Moreover, IIB-REs capacity to generate identically sized 2b-RAD tags (IIB-RE-dependent) across all samples [38,40] and cleave at both strands of DNA removes the need for a post-digestion fragment size selection step. These characteristics also prevent fragment size [41] and strand [42] sequencing bias which can compromise genotyping calls, as seen in other RADseq protocols. One disadvantage compared to other RADseq methods is that 2b-RAD may be inappropriate where accurate mapping against a highly duplicated/polyploid reference genome is required due to short fragment size production (33-36 bp) [43]. Finally, bias from PCR duplicates, sequencing errors and allele dropout can be introduced in all RADseq protocols.

In our study, we were able to rapidly and cost-effectively generate several hundred Single Nucleotide Polymorphism (SNP) markers for *R. ecuadoriensis* allowing for resolution of regional population genetic structure. Furthermore, by comparing the performance among the three IIB-REs, we were able to recommend the appropriate IIB-RE and read depth to employ in order to yield a given number of SNP markers for *R. ecuadoriensis* and, our implication, other members of the *Rhodnius* genus.

Methods

Sample collection and gDNA extraction

A total of 20 samples of *R. ecuadoriensis* were selected from the communities of La Extensa, Chaquizhca, and Coamine in the Province of Loja (Southern Andean region), and from the community of Bejuco in

Manabí (Central coast region) in Ecuador (Fig 1). Triatomines were captured in field surveys from June 2006 to July 2013 (Table 1). For each sample, head, legs and thoraxes were dissected and preserved in 100% alcohol. A modified salt extraction protocol of Aljanabi and Martinez [44] was used to extract total gDNA from *R. ecuadoriensis* heads, legs and thoraxes (hind gut excluded). The modified protocol involved an additional overnight chitinase digestion step, as well as one overnight 75% ethanol wash to ensure purity. gDNA concentrations and purity ratios assessments were obtained by using NanoDrop® ND-1000 Spectrophotometer (NanoDrop Technologies, Inc.). Integrity of the extracted DNA was evaluated by agarose electrophoresis and highly fragmented samples were excluded from subsequent analysis.

Type IIB restriction enzymes selection by *in silico* digestion

Initial selection of potential IIB-REs for our 2b-RAD protocol involved an *in silico* digestion of the *R. prolixus* genome, which is available from Genbank (accession code: KQ034056.1). For this purpose, 7 REs (*Alf*I, *CspC*I, *BsaX*I, *Sbf*I, *EcoR*I, *Bcg*I and *Kpn*I) were screened. Three type IIB RE, namely *Alf*I, *Bcg*I and *CspC*I were chosen based on the total number of restriction fragments produced *in silico* based on the draft *R. prolixus* genome (www.vectorbase.org), financial resources, known effectivity in previous studies [31-33,38] and authors previous experience working with those enzymes [45]. We expected REs with abundant *in silico* restriction sites to show larger coverage variability among samples, at lower read depths. On the contrary, REs with less abundant restriction sites *in silico* could provide more exploitable markers at lower read depths.

2b-RAD library preparation and Illumina sequencing

Libraries were prepared using the 2b-RAD protocol proposed by Wang *et al.* [38] (Box 1 and Fig 2). Reaction mix and PCR conditions varied (S2 Table) depending on which IIB-REs was used. First, approximately 100-400 ng of high quality gDNA from each sample were digested separately by each IIB-REs, producing IIB-REs-specific uniform length fragments (32 bp, 35 bp and 33 bp for *Alf*I, *Bcg*I and *CspC*I, respectively) with random overhangs. To confirm that the restriction reaction took place appropriately, equal amounts of digested DNA (dDNA) and gDNA from the same sample were visualized on a 1 % agarose gel. Subsequently, the dDNA of each sample was ligated to a pair of partially double-stranded adaptors with compatible and fully degenerated overhangs (5'NNN3'). Finally, the obtained 2b-RAD tags were amplified to introduce a sample-specific 7bp barcode and the Illumina NGS annealing sites using two different pairs of sequencing primers. A 1.8% agarose gel electrophoresis of the PCR products was performed to verify the presence of the expected 150 bp target band (fragment, barcodes and adaptors included). In order to ensure an approximately equimolar contribution of each sample to the library, the exact amount of each PCR product was measured from the intensity of the target band in a digital image of the 1.8% agarose gel. We prepared three libraries in total, one for each IIB-RE, according to the relative concentration of each sample. The purification of the libraries from high-molecular weight fragments and primer-dimers was achieved first by

removing the target band on agarose gel from each sample among the three libraries and eluting them in water overnight; followed by DNA capture with magnetic beads (SPRIselect® Beckman Coulter) based on the Solid-Phase Reversible Immobilisation method [46]. The DNA concentration in the purified libraries was quantified with a Qubit® Flourometer (Invitrogen) and the libraries were assembled in one single pool according to their relative concentrations. The library pool was sequenced on MiSeq (Illumina, San Diego, CA, USA) with a single 1x50 bp setup using 'Version2' chemistry at the Science for Life Laboratory (SciLifeLab, Stockholm, Sweden), which also implemented the reads demultiplexing and quality-filtering (Box 1 and Fig 2). Raw sequencing data can be accessed through Dryad digital repository (datadryad.org, DOI available on acceptance).

***In silico* assay to determine read depth vs locus recovery**

The quality of demultiplexed and quality-filtered raw reads was verified by using FastQC software [47]. Subsequently, custom-made Python scripts were used for trimming the adaptors and then filtering the reads on the IIB-RE-specific recognition site (Box 1 and Fig 2). For each of the three libraries (*Alf*I, *Bcg*I and *CspC*I) we sought to determine the relationship between sequencing effort (number of reads) and the total yield of polymorphic loci (set at up to two SNPs per locus). Therefore, we subsampled the total number of reads for each library in each individual using the fasta-subsample package from MEME SUITE [48] portal. This script randomly subsampled 25%, 50%, and 75% of total read in triplicate to assess variability. This process resulted in 10 datasets per IIB-RE library: nine representing the three subsampling repetitions of the fixed percentages and only one from the total (100%) reads itself.

To estimate the polymorphic loci growth rate among the three RE, a nonlinear least square fitting (NLS) approach [49,50] was used within the R software [51] package NLS [52]. Specifically, NLS algorithm fits to the data by approximating a nonlinear function to a linear one and in which an iterative process is required to calculate the optimal parameter values for the growth rate [49,50, 53]. Different built-in NLS models were tested in order to find the best fit to our data. These models were represented each with a different version of the Power-law equation [54]:

$$Y = aX^b$$

Where Y is the expected number of polymorphic loci at reads yield X; a is the starting amount expected when $X = 0$; b is the estimate of the relative change of Y in relation to a unit change in X. Detailed description of the equations used for each dataset is provided in S3 Table.

Genotype calling and filtering

All datasets created were analysed separately using STACKS software version 1.42 [55], in which *in silico* assembly of loci and individual genotyping was performed by running the DENOVO_MAP.PL pipeline (Box 1 and Fig 2). Due to the failure of the protocol in one of the samples from the *Alf*I library (likely as a result of

low gDNA quality), we decided to discard this sample from the other two datasets to avoid biased results in the *de novo* assembly. After several parameters adjustments, we set the minimum number of identical raw reads necessary to create a stack (-m) to 5; the default value for the numbers of mismatches allowed between loci when building a locus in a single individual, and when comparing across all individuals in order to build the population catalogue (-M and -n, respectively); the bounded SNP calling model for identifying a SNP and estimating the sequencing error rate for calling at that SNP (--bound) ranging from 0 to 0.05; and finally the significance level required to call a heterozygote or homozygote (--alpha) to 0.01. The EXPORT_SQL.PL script was used to export loci with the same polymorphism level (loci with up to 2 SNPs shared by at least the 80% and the 90% of samples) from the MySQL database for all datasets analysed in STACKS for each IIB-RE (Box 1 and Fig 2). Moreover, for all datasets the information exported represented the polymorphic loci with up to two SNPs with the sharing by at least the 80% and 90% of the samples.

Population genomic analysis

Although both total number of samples (N=19) and sample size per community (N=4-5) was low, we conducted pilot explorations of the population structure of *R. ecuadoriensis* in the study area. We retained polymorphic loci shared by at least 90% of the samples and characterised by the presence of 1 and 2 SNPs and minor allele frequency of 0.01. We performed preliminary genomic analysis using two different datasets: i) one containing 361 polymorphic loci obtained from 18 samples processed with the *Bcg*I IIB-RE (one sample was excluded from the analysis due to the high level of missing data); ii) the second containing 1225 polymorphic loci obtained from 19 samples processed with the *CspC*I IIB-RE. The number of markers obtained for the *Alf*I dataset derived from digestion with *Alf*I was too low to be used for the preliminary assessment of genomic structure of this particular sample. When two SNPs were recovered at a locus, only the first was retained for the analysis to exclude tightly linked SNP variation.

ARLEQUIN version 3.5 [59] was used to calculate non-hierchical analysis of molecular variance [AMOVA; 60]. To deal with missing data, the locus-by-locus option was set. Bayesian clustering implemented in STRUCTURE 2.3.4 [61] was conducted to investigate the most likely number of clusters of genetically related individuals excluding the locality origin (model LOPRIORI). After several trials, a burn-in of 300 000 followed by 3 million runs for K=1 to K=4 and 5 iterations per each K value was set; admixture model and correlated allelic frequencies were assumed. The most probable number of clusters was identified on the delta K implemented online with STRUCTURE HARVESTER [62]. Then, in order to confirm our polymorphic loci was *Rhodnius sp.*-related, we also aligned the total polymorphic loci shared by at least the 90% of samples obtained from *Bcg*I and *CspC*I datasets onto the reference *R. prolixus* genome using BOWTIE 1 [63]. The highest alignment score (*--best*) was chosen and no more than 3 mismatches (*-v*) were allowed.

Box 1. 2b-RADseq library and data preparation for upstream analysis.

1. gDNA extraction. gDNA is extracted from crushed thorax, legs and head of Triatomines. Etter *et al*. [27] provided useful recommendations on gDNA quality and concentration for optimal restriction enzyme digestion.

2. Library preparation. All RADseq protocols possess similar steps (Fig 2), however, technical particularities (see [40,56]) exist within each stage such as digestion using type IIB restriction enzymes and ligation of costum-made adaptors in 2b-RAD protocol.

3. Sequencing. As most of techniques, 2b-RAD tags can be sequenced on Illumina platforms which provide a range of sequencing lengths (50-300 bp) along with other options such as single (forward) or paired-end sequencing (forward and reverse reads) [40].

4. Raw data cleaning. Before genotyping, quality of raw 2b-RAD reads is assessed using different software such as FastQC (see [47] manual). Then, custom-made scripts are used to remove adaptors (trimming) and to keep (filtering) quality reads with the expected forward and reverse IIB-RE-specific recognition site. Finally, given that STACKS pipeline needs all reads in one direction, reads with reverse recognition site need to be forwarded (forwarding). Following these principles, scripts can be coded using Python and PERL programming languages, or alternatively, the PROCESS_RADTAGS module of STACKS can be used for the same process.

5. Genotype calling. Several software options [40,43] are available for SNP calling from RADseq data. In STACKS [55], the algorithm, first, reconstructs stacks (alleles exactly matching reads of each sample (-m); then, either merges them as a single polymorphic locus or keep them as a monomorphic locus depending on the number of nucleotide mismatches (-M), or if the stacks have repetitive sequences, they will be removed from the pipeline; finally, each sample information is stored in a catalog (stored in the MySQL repository) containing the consensus of all loci and alleles in the entire population (See [55] tutorials). There is no rule of thumb when setting SNP calling parameters, and sensitivity analysis are always recommended to target the best values. Perhaps, one should aim for detecting the highest number of SNPs while keeping a low error rate, which is controlled mainly by the -m, -M, and -n parameters. Moreover, the -bound and -alpha parameters deal with true heterozygote/homozygote calling errors at a determinate locus.

6. Genotype filtering. The EXPORT_SQL.PL program allows exporting the STACKS data in a compact format (TSV or XLS) and specify different filters such as loci with a determinate number of SNPs, alleles per locus and a percentage of sharing among samples. At this stage, additional filtering is recommended to remove loci or samples with a large amount of missing data and set a threshold for the polymorphic sharing by a fixed number of samples.

7. Analysis. Finally, the POPULATIONS program in STACKS exports previously identified loci in a standard output format such as GENEPOP or STRUCTURE. From this stage, the biological information is ready to use for further conventional population genetics analysis (AMOVA, and Bayesian clustering) or more recent approaches such as landscape genetics [57,58].

Results

gDNA extraction and *in silico* digestion

The extraction method allowed us to obtain RNA-free genomic DNA from all twenty samples with an average DNA concentration (ng/µL) of 62.77 ± 33.75 (s.d.); and average DNA purity ratios of 1.81 ± 0.05 (s.d.) and 1.81 ± 0.62 (s.d.), for absorbance at 280/260 and at 260/230, respectively (see S1 Table for detailed information). The *in silico* digestion on *R. prolixus* genome sequence by *Alf*I, *Bcg*I and *CspC*I IIB-REs produced 204895, 103268 and 69984 putative cut sites, respectively.

2b-RAD protocol

The 2b-RAD experimental approach used in this study was effective for *R. ecuadoriensis* gDNA samples using any of the three IIB-REs (Fig 3), except for one sample (ID: CQ12, see Table 1) digested by *AlfI*, which was not included to the pool for sequencing. A fifty-nine 2b-RAD pool was made of nineteen samples digested by *Alf*I, twenty by *Bcg*I, and twenty by *CspC*I IIB-REs.

Sequencing data filtering and *de novo* analysis

The Illumina NGS yielded a total of 14.8 million de-multiplexed and filtered-quality reads, approximately 3, 6.2 and 5.6 million reads (Mreads) for *Alf*I, *Bcg*I, and *CspC*I, respectively. FastQC analysis showed high per-base quality scores (> 32) for the reads of all samples processed with all the three IIB-REs. After trimming the adaptors and filtering the IIB-RE-specific recognition site, 2.9, 5.8 and 4.8 Mreads for *Alf*I, *Bcg*I, and *CspC*I respectively were retained (Fig 4). The average trimmed Mreads per sample for each IIB-REs was 0.15 ± 0.06, 0.30 ± 0.04 and 0.25 ± 0.07. The number of reads subsampled and the total polymorphic loci for each IIB-REs are reported in Table 2. STACKS reference genome free runs assembled and identified a catalogue of loci from each of the datasets. The EXPORT_SQL.PL script was used to extract two datasets which included all the polymorphic loci with up to 2 SNPs shared by at least the 80% and the 90% of samples from each of the set percentages (25%, 50%, 75%, 100%) among the three replicates. We found only minor variation in the number of polymorphic loci called for each of the three subsampling replicates in all IIB-REs libraries. The average number of exported polymorphic loci obtained among replicates and from the total number of reads for each IIB-REs is reported in Table 2.

We observed growth in the number of loci recovered as we increased the read depth for all enzymes (Fig 5). However, while increasing read depth led to corresponding medium and minor gains in locus number for *Bcg*I and *Alf*I, respectively; for *CspC*I this number of loci is highlighted by a greater exponential growth in comparison to the other RE. Our results of best fit model analysis (S3 Table) for each RE dataset were obtained by assessing different NLS models residual standard error, parameters significant p-values, number of iterations to convergence, the correlation between y and predicted values, and Akaike Information Criterion (AIC). In the first dataset (Fig 5A), we found that logarithmic ($y \sim a + b\ln(x)$), geometric ($y \sim ax^{bx}$) and exponential ($y \sim ae^{(bx)}$) NLS equations best fit to the *AlfI*, *BcgI* and *CspCI* datasets, respectively, allowing the estimation of growth rate parameters α and b (S3 Table). As for the second dataset (Fig 5B), geometric ($y \sim ax^{bx}$) and Power-law ($y \sim ax^{b}$) equations converged the best fit and parameters estimation for *AlfI* and *BcgI*, and *CspCI*, respectively (S3 Table). Detailed statistical analysis is provided in S1 Code.

Preliminary population genomics analysis

The non-hierchical AMOVA carried out on all the four community samples for both datasets (*Bcg*I and *CspC*I) detected a strong signal of genetic structuring across the study area, with high statistically significant (P< 0.0001) global F_{ST} values of 0.20452 (*Bcg*I) and 0.39327 (*CspC*I). The most likely number of genetic clusters (K) identified by STRUCTURE was 2 for both datasets: on one side the 3 samples from Loja region (CE, EX, CQ) were grouped together, on the other side the sample from Manabí (BJ) was considered as a distinct cluster (Fig 6). The alignment to the *R. prolixus* reference genome resulted in a 42% and 31% of polymorphic loci aligned for *BcgI* and *CspCI*, respectively, likely as an expected genomic variability between

the *R. ecuadoriensis* and the available R. prolixus reference genome as well as the difficulty in mapping short reads.

Discussion

Our data demonstrate the power of 2b-RAD as a valid genotyping approach that can be applied to CD vectors for which either no reference genome exists or, as in our case, a reference genome exists within the same genus. Our data broadly support the assertion of Wang et al. [38] that the 2b-RAD approach provides a simple, cost-effective and robust means of generating genome wide SNP data for non-model organisms. In our experiment, library preparation and sequencing was completed within a month and the cost per sample was approximately $18 USD (library preparation and sequencing cost), as compared to $30 USD per sample in other RADseq methods [39]. In fact, costs and technical complexity are two of the key factors when considering a RADseq protocol for a determinate study [56].

Type IIB restriction enzymes performance

In our study, we have gone somewhat further than a proof-of-principle by evaluating the performance of three distinct 2b-RAD restriction enzymes, pre-screened *in silico* for their performance in terms of marker density against the *Rhodnius prolixus* genome [23]. Our methodological development aim was to test the predictability of the *in silico* cutter and to provide recommendations for suitable read depths, marker numbers and sample sizes for studies involving *Rhodnius sp.* vectors. We expected that an abundant *in silico* enzyme cutter would provide less usable molecular markers at lower read depths (Fig 4). It is important to highlight that, enzyme performance *in silico* in terms of number of restriction sites is not necessarily the same in an actual experiment due to genome size, nucleotides distribution, depth of coverage and GC composition [27,40,64]. Thus, a pilot experiment always offers valuable information on actual restriction enzyme performance.

Random re-sampling (rarefaction) of our datasets revealed distinct relationships between read depth and marker (polymorphic locus) number between the different enzymes, *CspC*I*, Bcg*I and *Alf*I (Fig 5) broadly in line with predictions of the number of usable markers (Fig 4). As such *CspC*I produced the largest amount of polymorphic markers whatever the read depth, thus, evidencing its experimental performance in *R. ecuadoriensis* and likely to other *Rhodnius* sp. vectors. *Alf*I and *Bcg*I on the other hand showed a marked tendency of deceleration for marker recovery as read depth increases. However, *Alf*I does show the initially steeper growth, in line with predictions that *Alf*I cut sites in the *R. prolixus* genome are more abundant (*Alf*I= 204895 sites, *Bcg*I= 103268 sites and *CspC*I= 69984 sites). Additionally, we were able fit nonlinear regression models to the data and estimate growth rate parameters for each enzyme (Fig 5). Although the model function varies per enzyme and dataset, all of them follow an exponential growth pattern which is more evident in *CspCI* datasets. Although the model function applied to the second *CspCI* dataset (Fig 5B)

did not entirely fit the data, it constitutes the best fit compared to generalized linear models or more complex NLS fitting functions. Fitting NLS models to fewer data points for parameter estimation is challenging, however, based on our best-fit selection process we were confident that by substituting x for a determinate read depth we can obtain an estimate of polymorphic loci growth per restriction enzyme. We hope this will be helpful to others planning similar studies.

At the read depth, we achieved on one Illumina MiSeq single-ended run across 20 *R. ecuadoriensis* DNA samples we generated 1244 markers for *CspC*I, 367 for *Bcg*I and 68 for *Alf*I. Even the lowest of these values eclipses the size of marker panels currently in use to explore Triatomine population genetics [8,13-20]. However, to generate read depths to exploit the higher density RE cutters (e.g. *Alf*I, *Bcg*I) a HiSeq approach might be more sensible. On the other hand, based on our data, *CspC*I can be expected to generate the best coverage and over a thousand polymorphic markers for approximately sixty vector samples on one MiSeq run. Interestingly, Graham *et al.* [65] experiments assessed the impact of degraded gDNA in a modified double-digest-RAD protocol [37] in MiSeq platform and found a significant correlation between degradation, and read low quality and read loss. They also suggested that a higher throughput platform, HiSeq, and short fragment producer protocols, such as 2b-RAD, could help dealing with degraded gDNA and subsequent sequencing problems. As such, 2b-RAD might be an option for research teams with large and long-term stored triatomine bug collections, in which gDNA might already started a degradation processes. In our study, *CspCI* is the best candidate for generating enough usable markers, seconded by *BcgI*, and is likely that a sequencing platform such as HiSeq can exploit a higher number of markers for both enzymes.

Preliminary population genomic analysis

As well as 'range finding' for the application of 2b-RAD sequencing to triatomine populations, our second aim was to undertake preliminary population genomic analysis to explore genetic structuring in our study region. To this end, we focused on datasets generated with *Bcg*I and *CspC*I since they presented higher numbers of polymorphic loci. An AMOVA indicated a significant proportion of variation was explained by between-population differences for both datasets. Moreover, we demonstrated the feasibility of our markers to distinguish signal of structuring among populations in both *BcgI* and *CspCI* datasets. By using a Bayesian clustering framework our markers from both data sets detected two distinct clusters without previous location information, one of them was Bejuco, the clear geographic outlier with respect to Loja populations. Morphometric studies of *R. ecuadoriensis* in Ecuador would also predict a similar pattern of diversification [66]. However, inter-population diversification in Loja might be happening (Villacís *et al.* Unpublished work) at a rate that could not be detectable by isolation by distance hypothesis or conventional population analysis techniques. Our genomic information coupled with a landscape genetics/genomics framework could test whether landscape heterogeneity and environmental variable are driving such process [58].

Overcoming 2b-RAD pitfalls and study limitations

Earlier in the manuscript we presented the notion that, fewer steps, simplicity, cost-effectiveness, fragment size and strand bias absence, are advantages of using a 2b-RAD protocol compared to other RADseq methods. Nevertheless, researchers must be aware of potential pitfalls and sources of bias accompanying all RADseq protocols, as well as most NGS-based methods. Development of analysis and software tools to deal with this type of issues during NGS is an active and evolving research field [67]. During the initial steps of library preparation, degraded gDNA seems to have an impact in read quantity and quality in all other RADseq protocols than in 2b-RAD [65], however, guidelines [27] for assessing gDNA quality should be implemented in all protocols. Another drawback in all RADseq methods is that polymorphism can occur at the restriction site, defined as allele dropout (ADO), preventing enzymes to cut that location which makes impossible to recover that SNP allele (null allele) [40,68]. ADO will have a direct impact in the estimation of allele frequencies and consequently in overestimation/underestimation of F-statistics as individual heterozygote at the null allele will be recognised as homozygote. However, filtering loci successfully genotyped among a high percentage of the samples can help to remediate the problem [40]. PCR duplicates arise in all RADseq protocols with a PCR step, and only identifiable in protocols with a random shearing digestion (original RADseq protocol [35,36]) as duplicates fragments are identified by having same length. Another promising approach described by Andrews *et al.* [40] to identified PCR duplicates is using degenerated base regions within sequencing adaptors to mark parent fragments. However, Puritz *et al.* [43] highlighted that, though untested, skewed allele frequencies by PCR artefacts have little effect in statistical bias within loci and thereby genotype calling errors. No less important are sequencing errors introduced in all Illumina instruments. Although several genotype-calling algorithms account for sequencing errors, a high depth sequencing coverage ($\geq$ 20x) is always recommended. Finally, sequencing depth variability among loci could reduce genotyping accuracy for some less covered loci requiring multiplexing less individuals per sequencing lane at an increased cost [40,56].

In our study, most of the above issues encountered in RADseq have been circumvented either during the library preparation or the raw data filtering. Nevertheless, our main challenge is the absence of a reference genome to which map short reads in order to ensure that all markers do indeed belong to *R. ecuadoriensis* and not to microorganisms such as bacteria and fungi. Furthermore, it may be important to differentiate between mitochondrial and autosomal loci or sex-specific chromosomes that might have an effect in population divergence analysis. To overcome this difficulty, we adopted a stringent approach during raw data trimming, genotype calling, filtering loci shared by a high proportion of the samples and removal of loci and samples with high amount of missing data.

Further applications

Landscape genetics/genomics is a powerful and relatively new approach to explore the underlying spatial processes that affect genetic diversity in biological organisms [57]. Isolation-by-distance and isolation-by-

resistance are common null hypothesis in landscape genetics, where more complex ecological and environmental processes might be at play. This framework tool box such as causal modelling and environmental association analysis have the potential [57,58, 69-71] to undercover whether the same is likely to be the case for *R. ecuadoriensis* genetic structuring and dispersal in Ecuador. In our study, the main limitation to carry out a wide range of conventional between and within population analysis was the sample size per population; instead, we performed analysis of the entire study area.

The high-resolution genotyping approach we have developed in this study now paves the way for landscape genetic/genomics analysis in vector-parasite systems [58], with genuine potential insights for rational disease and entomological control. For example, landscape genetics approaches expanded our understanding of the natural and human-aided dispersal dynamics of the invasive Asian tiger mosquito, *Aedes albopictus* [72]. Similarly, insecticide resistance gene spread in *Anophles sinenis* has been tracked in China using landscape genetics approaches, demonstrating multiple origins and the importance of long term agricultural insecticide usage [73]. More widely, high resolution SNP datasets are increasingly in use to explore the local and international spread of important disease vectors e.g. [29,74].

2b-RAD typing not only promises a potential applicability for population genetic studies but also for linkage and quantitative loci mapping given that marker density can be controlled using selective adaptors [38]. In fact, the STACKS package through its GENOTYPE pipeline could allow the construction of genetic maps from F_2 or backcrosses of *R. ecuadoriensis* or other triatomine species.

In conclusion, the decreasing cost and increasingly simplicity of approaches to generate high resolution SNP data puts such tools increasingly in the hands of researchers in endemic countries working on non-model organisms that act as the vectors of Neglected Tropical Diseases. Furthermore, the analytical framework, incorporating detailed spatial and environmental variation into genetic analyses, is now in place to facilitate a better understanding of the biology and dispersal of disease vectors.

Acknowledgements

We are thankful to the following people for their advice and help: Babbucci M (University of Padua, Italy); Schwabl P (University of Glasgow, UK); Dario MA (Fundação Oswaldo Cruz, Brazil); Flores M (Fiocruz Rondônia, Brazil).

Author Contributions

Conceptualization: Hernandez-Castro LE, Paterno M, Grijalva MJ, Llewellyn MS.

Data curation: Hernandez-Castro LE, Paterno M, Llewellyn MS.

Formal analysis: Hernandez-Castro LE, Paterno M, Llewellyn MS.

Funding acquisition: Hernandez-Castro LE, Paterno M, Grijalva MJ, Llewellyn MS.

Investigation: Hernandez-Castro LE, De Noia M, Llewellyn MS.

Methodology: Hernandez-Castro LE, Paterno M, Llewellyn MS.

Project administration: Hernandez-Castro LE, Grijalva MJ, Llewellyn MS.

Resources: Villacís AG, Andersson B, Costales JA, Ocaña-Mayorga S, Yumiseva CA.

Software: Hernandez-Castro LE, Paterno M, Llewellyn MS.

Supervision: Grijalva MJ, Llewellyn MS.

Visualization: Hernandez-Castro LE, Paterno M.

Writing – original draft: Hernandez-Castro LE, Paterno M.

Writing – review & editing: Hernandez-Castro LE, Paterno M, G VillacísA, Andersson B, Costales JA, De Noia M, Ocaña-Mayorga S, Yumiseva CA, Grijalva MJ, Llewellyn MS.

References

1. Dias J, Silveira A, Schofield C. The impact of Chagas disease control in Latin America: a review. Mem Inst Oswaldo Cruz. 2002;97(5):603–12. pmid:12219120.

2. Moncayo Á, Silveira AC. Current epidemiological trends for Chagas disease in Latin America and future challenges in epidemiology, surveillance and health policy. Mem Inst Oswaldo Cruz. 2009;104:17–30. pmid:19753454.

3. Pinazo M-J, Gascon J. The importance of the multidisciplinary approach to deal with the new epidemiological scenario of Chagas disease (global health). Acta Trop. 2015;151:16–20. doi:10.1016/j.actatropica.2015.06.013. pmid:26187358.

4. Noireau F, Cortez M, Monteiro F, Jansen A, Torrico F. Can wild foci in Bolivia jeopardize Chagas disease control efforts? Trends Parasitol. 2005;21(1):7–10. doi:10.1016/j.pt.2004.10.007.

5. Schofield CJ, Jannin J, Salvatella R. The future of Chagas disease control. Trends Parasitol. 2006;22(12):583–8. doi:10.1016/j.pt.2006.09.011. pmid:17049308.

6. Tarleton RL, Reithinger R, Urbina JA, Kitron U, Gürtler RE. The challenges of Chagas Disease - grim outlook or glimmer of hope? PLoS Med. 2007;4(12):e332. doi:10.1371/journal.pmed.0040332. pmid:18162039.

7. Guhl F, Pinto N, Aguilera G. Sylvatic triatominae: a new challenge in vector control transmission. Mem Inst Oswaldo Cruz. 2009;104:71–5. pmid:19753461.

8. Ceballos LA, Piccinali R V., Marcet PL, Vazquez-Prokopec GM, Cardinal MV, Schachter-Broide J, et al. Hidden sylvatic foci of the main vector of chagas disease *Triatoma infestans*: threats to the vector elimination campaign? PLoS Negl Trop Dis. 2011;5(10):e1365. doi:10.1371/journal.pntd.0001365. pmid:22039559.

9. Stevens L, Dorn PL, Schmidt JO, Klotz JH, Lucero D, Klotz SA. Kissing Bugs. The vectors of Chagas. Advances in Parasitology. 2011;169–92. doi:10.1016/B978-0-12-385863-4.00008-3.

10. Lent H, Wygodzinsky PW. Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. Bulletin of the AMNH. 1979;163(3):125-520.

11. Grijalva MJ, Villacis AG. Presence of *Rhodnius ecuadoriensis* in sylvatic habitats in the southern highlands (Loja Province) of Ecuador. J Med Entomol. 2009;46(3):708–11. doi:10.1603/033.046.0339. pmid:19496445.

12. Grijalva MJ, Suarez-Davalos V, Villacis AG, Ocaña-Mayorga S, Dangles O. Ecological factors related to the widespread distribution of sylvatic *Rhodnius ecuadoriensis* populations in southern Ecuador. Parasit Vectors. 2012;5:17. doi:10.1186/1756-3305-5-17.

13. Fitzpatrick S, Feliciangeli MD, Sanchez-Martin MJ, Monteiro FA, Miles MA. Molecular genetics reveal that silvatic *Rhodnius prolixus* do colonise rural houses. PLoS Negl Trop Dis. 2008;2(4):e210. doi:10.1371/journal.pntd.0000210.

14. Gourbière S, Dorn P, Tripet F, Dumonteil E. Genetics and evolution of triatomines: from phylogeny to vector control. Heredity. 2012;108(3):190–202. doi:10.1038/hdy.2011.71.

15. Brenière SF, Waleckx E, Magallón-Gastélum E, Bosseno M-F, Hardy X, Ndo C, et al. Population genetic structure of *Meccus longipennis* (Hemiptera, Reduviidae, Triatominae), vector of Chagas disease in West Mexico. Infect Genet Evol. 2012;12(2):254–62. doi:10.1016/j.meegid.2011.11.003.

16. García BA, de Rosas ARP, Blariza MJ, Grosso CG, Fernández CJ, Stroppa MM. Molecular Population Genetics and Evolution of the Chagas' Disease Vector *Triatoma infestans* (Hemiptera: Reduviidae). Curr Genomics. 2013;14(5):316–23. doi:10.2174/13892029113149990006.

17. Belisário CJ, Pessoa GCD, dos Santos PF, Dias LS, Rosa ACL, Diotaiuti L. Markers for the population genetics studies of *Triatoma sordida* (Hemiptera: Reduviidae). Parasit Vectors. 2015;8(1):269. doi:10.1186/s13071-015-0879-1.

18. Piccinali RV, Gürtler RE. Fine-scale genetic structure of *Triatoma infestans* in the Argentine Chaco. Infect Genet Evol. 2015;34:143–52. doi:10.1016/j.meegid.2015.05.030.

19. Stevens L, Monroy MC, Rodas AG, Hicks RM, Lucero DE, Lyons LA, et al. Migration and Gene Flow Among domestic populations of the Chagas insect vector *Triatoma dimidiata* (Hemiptera: Reduviidae) detected by microsatellite loci. J Med Entomol. 2015;52(3):419–28. doi:10.1093/jme/tjv002.

20. Almeida CE, Faucher L, Lavina M, Costa J, Harry M, Ferreira I, et al. molecular individual-based approach on *Triatoma brasiliensis*: inferences on triatomine foci, *Trypanosoma cruzi* natural infection prevalence, parasite diversity and feeding sources. PLoS Negl Trop Dis. 2016;10(2):e0004447. doi:10.1371/journal.pntd.0004447.

21. Stephenson JJ, Campbell MR, Hess JE, Kozfkay C, Matala AP, McPhee M V., et al. A centralized model for creating shared, standardized, microsatellite data that simplifies inter-laboratory collaboration. Conserv Genet. 2009;10(4):1145–9. doi:10.1007/s10592-008-9729-4.

22. Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MMG, et al. The revised *Trypanosoma cruzi* subspecific nomenclature: Rationale, epidemiological relevance and research applications. Infect Genet Evol. 2012;12(2):240–53. doi: 10.1016/j.meegid.2011.12.009.

23. Mesquita RD, Vionette-Amaral RJ, Lowenberger C, Rivera-Pomar R, Monteiro FA, Minx P, et al. Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. Proc Natl Acad Sci. 2015;112(48):14936–41. doi:10.1073/pnas.1506226112.

24. Avila ML, Tekiel V, Moretti G, Nicosia S, Bua J, Lammel EM, et al. Gene discovery in *Triatoma infestans*. Parasit Vectors. 2011;4(1):39. doi:10.1186/1756-3305-4-39.

25. Buarque DS, Braz GRC, Martins RM, Tanaka-Azevedo AM, Gomes CM, Oliveira FAA, et al. Differential expression profiles in the midgut of *Triatoma infestans* infected with *Trypanosoma cruzi*. PLoS One. 2013;8(5):e61203. doi:10.1371/journal.pone.0061203.

26. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. population genomics of parallel adaptation in threespine stickleback using sequenced RAD Tags. PLoS Genet. 2010;6(2):e1000862. doi:10.1371/journal.pgen.1000862.

27. Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. 2012;157–78. doi:10.1007/978-1-61779-228-1_9.

28. Guo Y, Yuan H, Fang D, Song L, Liu Y, Liu Y, et al. An improved 2b-RAD approach (I2b-RAD) offering genotyping tested by a rice (*Oryza sativa L*.) F2 population. BMC Genomics. 2014;15(1):956. doi:10.1186/1471-2164-15-956.

29. Rašić G, Filipović I, Weeks AR, Hoffmann AA, Bhatt S, Gething P, et al. Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, *Aedes aegypti*. BMC Genomics. 2014;15(1):275. doi:10.1186/1471-2164-15-275.

30. Evans BR, Gloria-Soria A, Hou L, McBride C, Bonizzoni M, Zhao H, et al. A multipurpose, high-throughput single-nucleotide polymorphism chip for the dengue and yellow fever mosquito, *Aedes aegypti*. Genes|Genomes|Genetics. 2015;5(5):711–8. doi:10.1534/g3.114.016196.

31. Barfield S, Aglyamova G V, Matz M V, Busss L, D'Amato F, Laux T, et al. Evolutionary origins of germline segregation in Metazoa: evidence for a germ stem cell lineage in the coral *Orbicella faveolata* (Cnidaria, Anthozoa). Proc Biol Sci. 2016;283(1822):1387–91. doi:10.1098/rspb.2015.2128.

32. Pauletto M, Carraro L, Babbucci M, Lucchini R, Bargelloni L, Cardazzo B. Extending RAD tag analysis to microbial ecology: a comparison between MultiLocus Sequence Typing and 2b-RAD to investigate *Listeria monocytogenes* genetic structure. Mol Ecol Resour. 2016;16(3):823–35. doi:10.1111/1755-0998.12495.

33. Pecoraro C, Babbucci M, Villamor A, Franch R, Papetti C, Leroy B, et al. Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*). Mar Genomics. 2016;25:43–8. doi:10.1016/j.margen.2015.12.002.

34. Manousaki T, Tsakogiannis A, Taggart JB, Palaiokostas C, Tsaparis D, Lagnel J, et al. Exploring a nonmodel teleost genome through rad sequencing--linkage mapping in Common Pandora, *Pagellus erythrinus* and comparative genomic analysis. Genes|Genomes|Genetics. 2016;6(3):509–19. doi:10.1534/g3.115.023432.

35. Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 2007;17(2):240–8. doi:10.1101/gr.5681207.

36. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 2008;3(10):e3376. doi:10.1371/journal.pone.0003376.

37. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 2012;7(5):e37135. doi:10.1371/journal.pone.0037135.

38. Wang S, Meyer E, McKay JK, Matz M V. 2b-RAD: a simple and flexible method for genome-wide genotyping. Nat Methods. 2012;9(8):808–10. doi:10.1038/nmeth.2023.

39. Toonen RJ, Puritz JB, Forsman ZH, Whitney JL, Fernandez-Silva I, Andrews KR, et al. ezRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ. PeerJ, Inc; 2013;1:e203. doi:10.7717/peerj.203.

40. Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016:81–92. doi:10.1038/nrg.2015.28.

41. Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. Special features of RAD Sequencing data: implications for genotyping. Mol Ecol. 2013;22(11):3151–64. doi:10.1111/mec.12084.

42. Guo Y, Li J, Li C-I, Long J, Samuels DC, Shyr Y. The effect of strand bias in Illumina short-read sequencing data. BMC Genomics. 2012;13(1):666. doi:10.1186/1471-2164-13-666.

43. Puritz JB, Matz M V., Toonen RJ, Weber JN, Bolnick DI, Bird CE. Demystifying the RAD fad. Mol Ecol. 2014;23(24):5937–42. doi:10.1111/mec.12965.

44. Aljanabi S, Martinez I. Universal and rapid salt-extraction of high quality genomic DNA for PCR- based techniques. Nucleic Acids Res. 1997;25(22):4692–3. doi:10.1093/nar/25.22.4692.

45. Paterno M, Schiavina M, Aglieri G, Ben Souissi J, Boscari E, Casagrandi R, et al. Population genomics meet Lagrangian simulations: Oceanographic patterns and long larval duration ensure connectivity among *Paracentrotus lividus* populations in the Adriatic and Ionian seas. Ecol Evol. 2017;00:1–17. doi:10.1002/ece3.2844.

46. DeAngelis MM, Wang DG, Hawkins TL. Solid-phase reversible immobilization for the isolation of PCR products. Nucleic Acids Res. 1995;23(22):4742–3. pmcid:PMC307455

47. Andrews S. Babraham Bioinformatics. FastQC: a quality control tool for high throughput sequence data. 2010. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

48. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37:W202-8. doi:10.1093/nar/gkp335.

49. Johnson ML. Nonlinear Least-Squares Fitting Methods. In: Methods in Cell Biology. 2008;84: 781–805. doi.10.1016/S0091-679X(07)84024-6.

50. Paine CET, Marthews TR, Vogt DR, Purves D, Rees M, Hector A, et al. How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. Methods Ecol Evol. 2012;3(2):245–56. doi:10.1111/j.2041-210X.2011.00155.x.

51. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.

52. Bates DM, Watts DG. Nonlinear Regression: Iterative Estimation and Linear Approximations. In: Nonlinear Regression Analysis and Its Applications. John Wiley & Sons, Inc.; 2008;32–66. doi:10.1002/9780470316757.ch2.

53. Brown AM. A step-by-step guide to non-linear regression analysis of experimental data using a Microsoft Excel spreadsheet. Comput Methods Programs Biomed. 2001;65(3):191–200. doi:10.1016/S0169-2607(00)00124-3.

54. Xiao X, White EP, Hooten MB, Durham SL. On the use of log-transformation vs. nonlinear regression for analyzing biological power laws. Ecology. Ecological Society of America. 2011;92(10):1887–94. doi:10.1890/11-0538.1.

55. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. NIH Public Access; 2013;22(11):3124–40. doi:10.1111/mec.12354.

56. Andrews KR, Hohenlohe PA, Miller MR, Hand BK, Seeb JE, Luikart G. Trade-offs and utility of alternative RADseq methods: Reply to Puritz et al. Mol Ecol. 2014;23(24):5943–6. doi:10.1111/mec.12964.

57. Manel S, Schwartz K, Luikart G, Taberlet P. Landscape genetics: Combining landscape ecology and population genetics. Trends in Ecology and Evolution. 2003;18:189-197. doi:10.1016/S0169-5347(03)00008-9.

58. Schwabl P, Llewellyn MS, Landguth EL, Andersson B, Kitron U, Costales JA, et al. Prediction and Prevention of Parasitic Diseases Using a Landscape Genomics Framework. Trends Parasitol. 2016; doi:10.1016/j.pt.2016.10.008.

59. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10(3):564–7. doi:10.1111/j.1755-0998.2010.02847.x.

60. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992;131(2):479–91. pmcid:PMC1205020.

61. Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. Genetics. 2000;155(2):945 LP-959. doi:10.1111/j.1471-8286.2007.01758.x.

62. Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4(2):359–61. doi:10.1007/s12686-011-9548-7.

63. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25. doi:10.1186/gb-2009-10-3-r25.

64. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12(7):499–510. doi:10.1038/nrg3012.

65. Graham CF, Glenn TC, McArthur AG, Boreham DR, Kieran T, Lance S, et al. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). Mol Ecol Resour. 2015;15(6):1304–15. doi:10.1111/1755-0998.12404.

66. Villacís AG, Grijalva MJ, Catalá SS. Phenotypic Variability of *Rhodnius ecuadoriensis* Populations at the Ecuadorian Central and Southern Andean Region. J Med Entomol. 2010;47(6):1034–43. doi:10.1603/ME10053.

67. Ekblom R, Galindo J. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity (Edinb). 2011;107(1):1–15. doi:10.1038/hdy.2010.152.

68. Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué C, Pudlo P, et al. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Mol Ecol. 2013;22(11):3165–78. doi:10.1111/mec.12089.

69. Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP. Landscape genetics: where are we now? Molecular Ecology. 2010;19(17):3496–3514. doi:10.1111/j.1365-294X.2010.04691.x.

70. Manel S, Holderegger R. Ten years of landscape genetics. Trends in Ecol & Evol. 2013;28(10);614-621. doi:j.tree.2013.05.012.

71. Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. A practical guide to environmental association analysis in landscape genomics. Mol Ecol. 2015;24(17):4348–70. doi:10.1111/mec.13322.

72. Medley KM, Jenkins DG, Hoffman EA. Human-aided and natural dispersal drive gene flow across the range of an invasive mosquito. Mol Ecol. 2014;24(2):284-295. doi:10.1111/mec.12925.

73. Chang X, Zhong D, Lo E, Fang Q, Bonizzoni M, Wang X, Lee MC, Zhou G, Zhu G, Qin Q, Chen X, Cui L, Yan G. Landscape genetic structure and evolutionary genetics of insecticide resistance gene mutations in *Anopheles sinensis*. Par & Vect. 2016;9:228. doi:10.1186/s13071-016-1513-6.

74. Brown JE1, Evans BR, Zheng W, Obas V, Barrera-Martinez L, Egizi A, Zhao H, Caccone A, Powell JR. Human impacts have shaped historical and recent evolution in *Aedes aegypt*i, the dengue and yellow fever mosquito. Evol. 2014;68(2):514-25. doi:10.1111/evo.12281.

Tables and Figures

Table 1. Information of total number of samples collected from Loja and Manabí, Ecuador used in this study. For each sample, the following information is reported: province, community, community ID, geographical coordinates, collection date and samples ID. Detailed sample information for each sample is provided in S1 Table.

Province	Community	Community ID	Coordinates latitude/longitude	Collection Date	Sample ID
Loja	Coamine	CE	-4.117625, -79.621013	Jun-10	CE7
				Jun-10	CE8
				Jun-10	CE9
				Jun-10	CE10
				Jun-10	CE11
Loja	La Extensa	EX	-4.043671, -79.359601	Jun-06	EX3
				Jun-06	EX8
				Aug-07	EX9
				Aug-07	EX10
				Aug-07	EX11
Loja	Chaquizhca	CQ	-4.231105, -79.594118	Jul-10	CQ4
				Jul-10	CQ10
				Jul-11	CQ11
				Jul-12	CQ12
				Jul-13	CQ13
Manabí	Bejuco	BJ	-0.9728, -80.3445	Jun-07	BJ5
				Jul-09	BJ10
				Jul-09	BJ11
				Oct-09	BJ12
				Oct-09	BJ13

Table 2. Relationship between reads and the number of polymorphic loci obtained from STACKS analysis. Mean values provided ± the standard error.

IIB-REs	% Subsampled and total reads	Reads (Mreads)	Polymorphic Loci – 1 and 2 SNPs	
			90% sharing	80% sharing
AlfI	25	0.7	28.7 ± 2.5	47 ± 3
	50	1.4	51 ± 2	75.3 ± 4
	75	2.2	57.3 ± 3.1	99 ± 1.7
	100	2.9	68	186
BcgI	25	1.5	50 ± 5.2	78.3 ± 3.1
	50	2.9	100.7 ± 6.4	149 ± 6.1
	75	4.4	162 ± 6.2	331 ± 10.4
	100	5.8	367	899
CspCI	25	1.2	46.3 ± 3.8	65 ± 6.9
	50	2.4	81.7 ± 2.1	154.3 ± 5.9
	75	3.6	341 ± 10.6	995 ± 23.4
	100	4.8	1244	2289

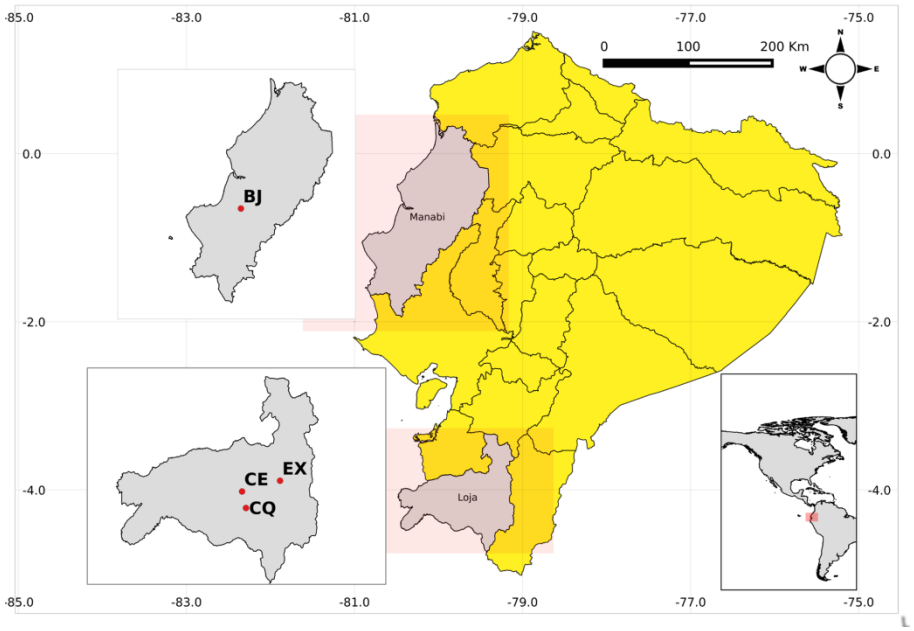


Figure 4. Map of the study area and the location of sampled communities in Ecuador. Red circles indicated the location of Coamine (CE), La Extensa (EX) and Chaquizhca (CQ) in the Province of Loja, and El Bejuco (BJ) in the Province of Manabí.

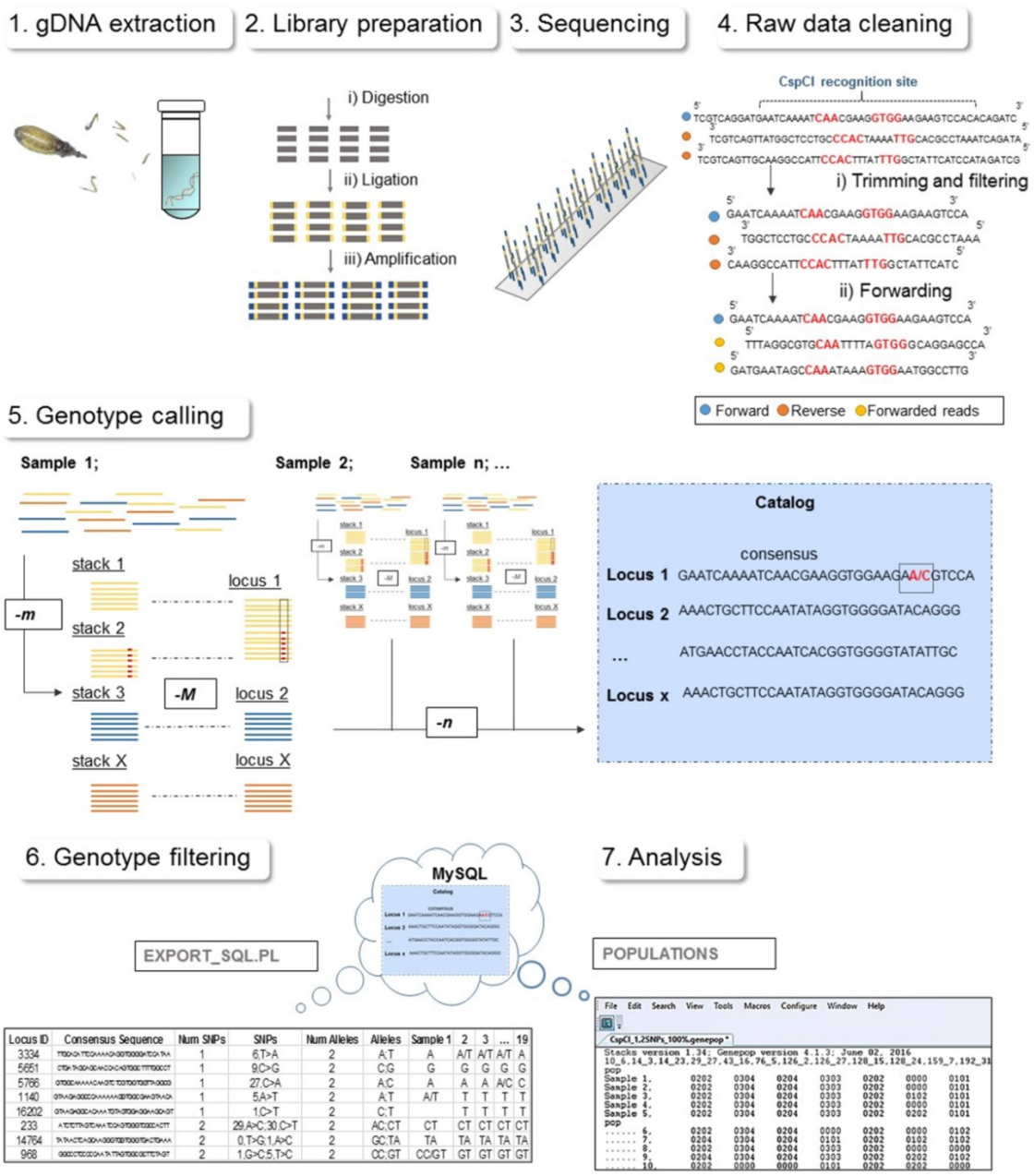


Figure 5. Step-by-step of 2b-RAD library and genomic data preparation for triatomines genomic population analysis. (1) gDNA is extracted from heads, legs and thorax of triatomine bugs. **(2)** After that, gDNA is processed using the 2B-RAD protocol [38] and **(3)** libraries are sequenced on Illumina instruments. **(4)** Once the data is delivered, it is trimmed and filtered before **(5)** used in genotyping software such as STACKS [55]. **(6)** Then, genotypes are exported from the cloud (MySQL repository) and filtered if large amount of missing data is present. **(7)** Finally, the polymorphic loci of interest are exported in conventional file formats for population genomic analysis. See Box 1 for details.

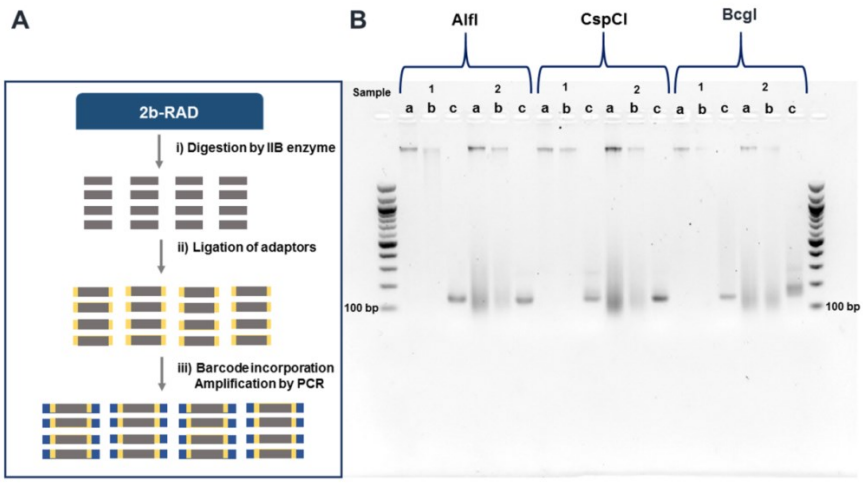


Figure 6. The 2b-RAD protocol. A) Description of the 2b-RAD library preparation. i) digestion by IIB-RE, ii) ligation to partially double-stranded adaptors with compatible and fully degenerated overhangs (5'NNN3'), iii) incorporation of barcodes and amplification by PCR as described in Wang *et al.* [36]. **B) Example of two samples processed with 2b-RAD protocol.** 1.8% agarose gel electrophoresis showing gDNA (a), digested DNA (b) and PCR product (c) in 2 samples for each of the 3 IIB-REs (*Alf*I, *Csp*CI and *Bcg*I).

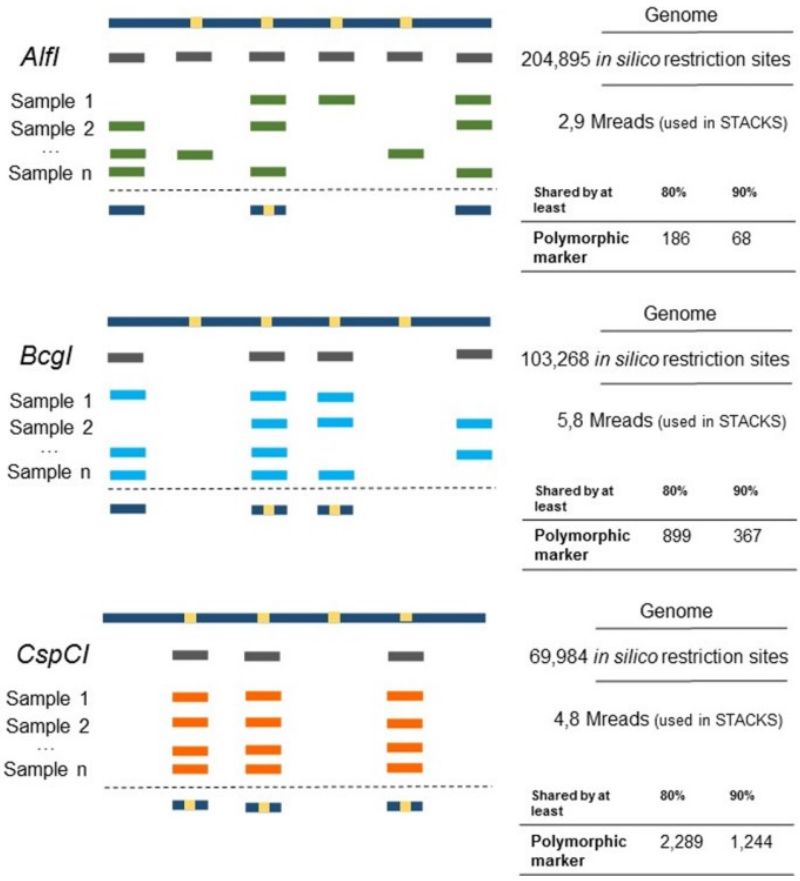


Figure 7. Comparison of read depth and marker identification in R. ecuadoriensis using 3 different type IIB restriction enzymes. In line with our *in silico* predictions, *AlfI*, an abundant *in silico* cutter did not produce enough molecular markers as compared to *BcgI* and *CspCI*, less abundant *in silico* cutters. In the diagram, enzymes with abundant *in silico* restriction sites (dark gray rectangles) within the genome (dark blue solid line with yellow squares or SNPs) are more likely to produce fragments (light blue, green and orange rectangles) at different location among samples during a random experiment, and therefore, this can result in not achieving enough read depth which can jeopardise polymorphic marker discovery (dark blue rectangles with a yellow square).

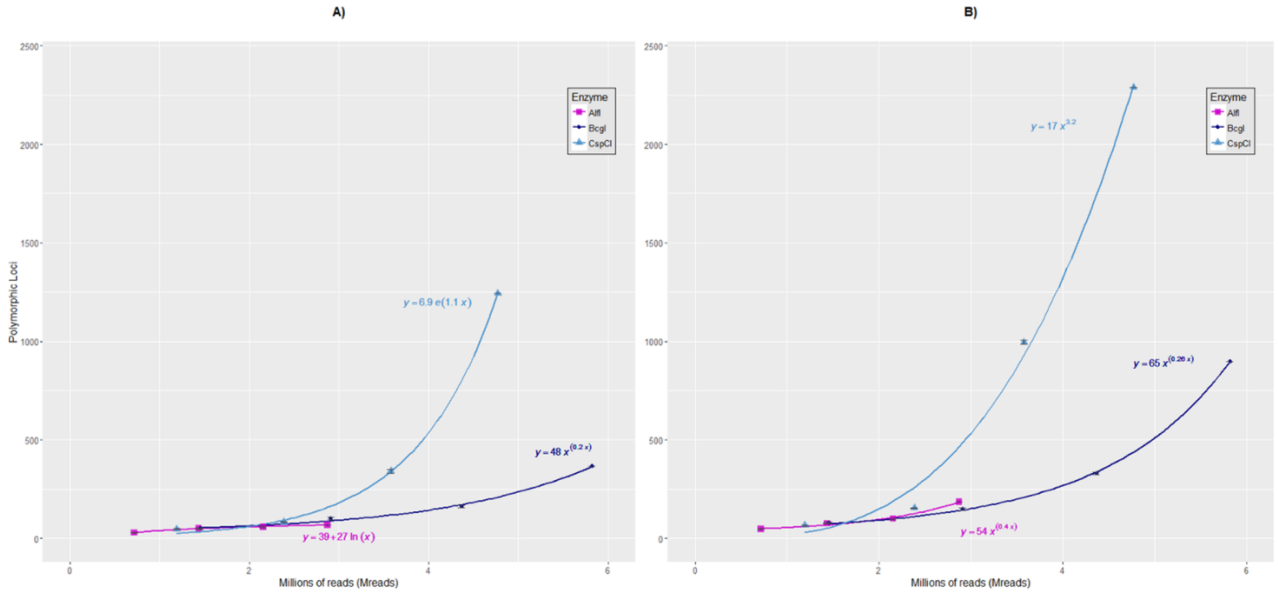


Figure 8. Relationship between the number of reads and the polymorphic loci obtained by each IIB-RE. Lines show the comparison of the relationship between the increased number of reads obtained by *Alf*I (Magenta square), *Bcg*I (Dark blue point) and *CspC*I (Light blue triangle) IIB-REs, and increasing numbers of polymorphic loci discovered after STACKS analysis. Different read abundances were obtained by randomly subsampling the dataset of each enzyme, and analysing in STACKS separately as independent datasets. In the figure **A)** Polymorphic loci with up to 2 SNPs shared by at least 90% of samples and best fit curves with resolved logarithmic (Magenta), geometric (Dark blue) and exponential (Light blue) equations. **B)** Polymorphic loci with up to 2 SNPs shared by at least 80% of samples and best fit curves with resolved geometric (Magenta and Dark blue) and Power-law (Light blue) equations.

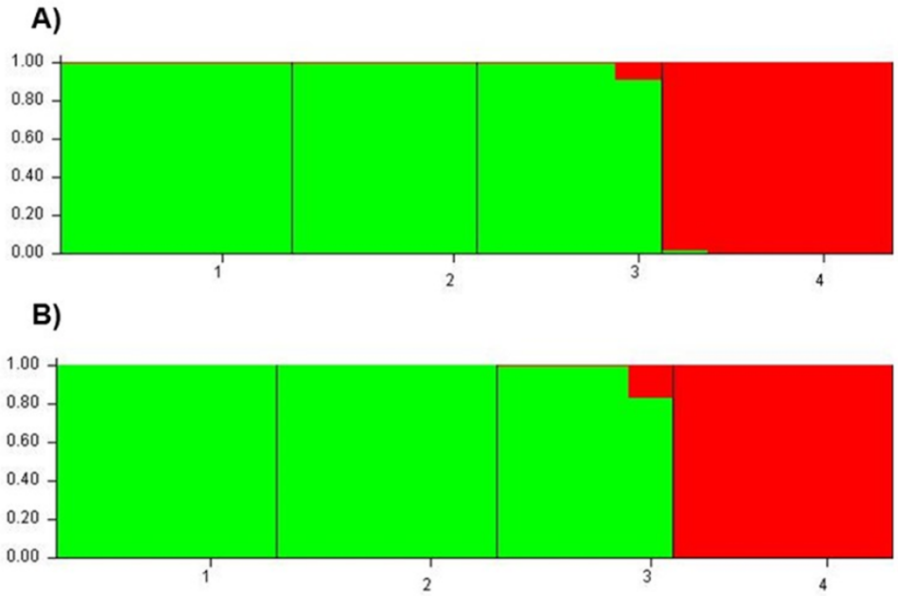


Figure 9. Genetic clusters (K = 2) assigned by STURCTURE. The green columns (1 - CE, 2 - EX and 3 - CQ) indicate the samples from Loja, and the red column (4 - BJ) the samples from Manabí. **A)** BcgI and **B)** CspCI datasets, respectively.

Supporting information

Table S1. Detailed information of *R. ecuadoriensis* samples used in this study.

Sample site and sample ID	*DNA concentration ng/µL*	*260/280 ratio*	*260/230 ratio*	*Latitude*	*Longitude*	*House ID*	*Habitat*	*Collection site*	*Collection date (DD/MM/YY)*	*Life stage*	*T. cruzi infection*
Chaquizhca, Loja.											
CQ4	37.2	1.82	1.59	-4.231105	-79.594118	CQ306	Intra-domicile	Kitchen wall	Fcol.09JUL2010	F	Positive
CQ10	66.4	1.95	1.63	-4.231105	-79.594118	CQ306	Intra-domicile	Kitchen wall	Fcol.09JUL2010	M	n.a
CQ11	43.8	1.85	1.57	-4.231105	-79.594118	CQ306	Intra-domicile	Kitchen wall	Fcol.09JUL2011	F	Negative
CQ12	32.9	1.86	1.39	-4.234881	-79.578047	CQ406	Peri-domicile	Chicken nest	Fcol.09JUL2012	M	Negative
CQ13	114.9	1.88	1.6	-4.12424	-79.7059	TC306	Peri-domicile	Chicken nest	Fcol.09JUL2013	M	Positive
Coamine, Loja.											
CE7	48.1	1.86	4.1	-4.117625	-79.62101394	CE509	Intra-domicile	Bed	Fcol.30JUN2010	M	Negative
CE8	74.8	1.76	1.83	-4.129398	-79.60071004	CE305	Peri-domicile	Chicken nest	Fcol.30JUN2010	F	Negative
CE9	49.1	1.82	2.55	-4.116617	-79.6163567	CE505	Peri-domicile	Hen house	Fcol.30JUN2010	M	Positive
CE10	97.3	1.88	2.18	-4.116617	-79.6163567	CE505	Peri-domicile	Hen house	Fcol.30JUN2010	F	Positive
CE11	81.4	1.83	1.75	-4.116617	-79.6163567	CE505	Peri-domicile	Hen house	Fcol.30JUN2010	M	Negative
La Extensa, Loja.											
EX3	29.5	1.73	1.47	-4.043671	-79.359601	EX608	Peri-domicile	Chicken nest	F. col26JUN2006	F	n.a
EX8	28.5	1.65	1.25	-4.043671	-79.359601	EX608	Peri-domicile	Bed	F. col26JUN2006	F	n.a
EX9	32.5	1.76	2.15	-4.04413	-79.35964	EX803	Peri-domicile	Chicken nest	F. col8AUG2007	F	Negative
EX10	52.7	1.74	1.4	-4.04413	-79.35964	EX803	Peri-domicile	Chicken nest	F. col8AUG2007	F	Negative
EX11	150.9	1.87	1.5	-4.04413	-79.35964	EX803	Peri-domicile	Chicken nest	F. col8AUG2007	F	Negative
Bejuco, Manabi.											
BJ5	72.7	1.81	1.64	-0.9728	-80.3445	BJN55-1	Peri-domicile	Rat nest	F. col23JUN2007	M	n.a
BJ10	25.7	1.67	1.73	-0.9728	-80.3445	TBJ1940	Peri-domicile	Chicken nest	Fcol.07JUL2009	M	n.a
BJ11	71.1	1.8	1.52	-0.9728	-80.3445	TBJ1941	Peri-domicile	Chicken nest	Fcol.07JUL2009	M	n.a
BJ12	40.5	1.86	1.93	-0.9728	-80.3445	TBJ2009	Peri-domicile	Chicken nest	Fcol.13OCT2009	M	n.a
BJ13	105.4	1.8	1.48	-0.9728	-80.3445	TBJ2010	Peri-domicile	Chicken nest	Fcol.13OCT2009	M	n.a

Table S2. Reagents and 2b-RAD protocol used in this study.

	1. Digestion	2. Ligation	3. Amplification
AlfI	**Reaction mix:** **Reagent | 1x** Buffer R 10X 0.6 µL SAM 100 µM 0.6 µL dd H_2O 0.05 µL AlfI * 2 U/µL 0.75 µL **Combine:** 2 µL Master mix + 4 µL genomic DNA **Reaction profile:** 37 °C for 2 hrs 65 °C for 20 min 16 °C for ∞	**Reaction mix:** **Reagent | 1x** dd H_2O 9.5 µL Buffer 2.5 µL ATP** 10 µM 0.5 µL Adaptor 2 2.5 µL Adaptor 3 2.5 µL T4 Ligase** 2.5 µL **Combine:** 20 µL Master mix + 5 µL digested DNA **Reaction profile:** 16 °C for 3 hrs 65 °C for 10 min 4 °C for overnight	**Reaction mix:** **Reagent | 1x** dd H_2O 14.7 µL Buffer 10 µL dNTP 2 mM 7.8 µL Primer F 2.5 µL pAMPF 1 µL pAMPR 1 µL Taq polymerase** 0.5 µL **Combine:** 37.5 µL Master mix + 10 µL ligated DNA + 2.5 µL barcoded-primer **Reaction profile:** 14 cycles: 98 °C for 5 sec 60 °C for 30 sec 72 °C for 5 sec Then: 72 °C for 5 min 16 °C for ∞
BcgI	**Reaction mix:** **Reagent | 1x** Buffer 0.8 µL SAM 100 µM 0.7 µL BcgI ** 2 U/ µL 0.5 µL **Combine:** 2 µL Master mix + 4 µL genomic DNA **Reaction profile:** 37 °C for 1 hr 65 °C for 20 min 4 °C for 2 hrs	**Reaction mix:** **Reagent | 1x** dd H_2O 9.5 µL Buffer 2.5 µL ATP** 10 mM 0.5 µL Adaptor 2 2.5 µL Adaptor 3 2.5 µL T4 Ligase** 2.5 µL **Combine:** 20 µL Master mix + 5 µL digested DNA **Reaction profile:** 16 °C for 2 hrs 65 °C for 10 min 4 °C for 2 hrs	**Reaction mix:** **Reagent | 1x** Same as *AlfI*'s reagent and quantities **Combine:** 37.5 µL Master mix + 10 µL ligated DNA + 2.5 µL barcoded-primer **Reaction profile:** 16 cycles: 98 °C for 5 sec 60 °C for 30 sec 72 °C for 5 sec Then: 72 °C for 5 min 16 °C for ∞
CspCI	**Reaction mix:** **Reagent | 1x** NEB Buffer 0.6 µL SAM 100 µM 1.2 µL CspCI ** 1.5 U/µL 0.75 µL **Combine:** **2** µL Master mix + 4 µL genomic DNA **Reaction profile:** 37 °C for overnight Refill samples with 0.2 µL CspCI after 2 hrs 65 °C for 10 min 16 °C for ∞	**Reaction mix:** **Reagent | 1x** dd H_2O 9.5 µL Buffer 2.5 µL ATP** 10 µM 0.5 µL Adaptor 2 2.5 µL Adaptor 3 2.5 µL T4 Ligase** 2.5 µL **Combine:** 20 µL Master mix + 5 µL digested DNA **Reaction profile:** 16 °C for overnight 16 °C for 10 min 4 °C for ∞	**Reaction mix:** **Reagent | 1x** Same as *AlfI*'s reagents and quantities **Combine:** 37.5 µL Master mix + 10 µL ligated DNA + 2.5 µL barcoded-primer **Reaction profile:** 16 cycles: 98 °C for 5 sec 60 °C for 30 sec 72 °C for 5 sec Then: 72 °C for 5 min 16 °C for ∞

Note: Adaptor 2 was made from oligonucleotide anti-ILL and 5ILL-NN; Adaptor 3 was made of oligonucleotide anti-ILL and 3ILL-NN from Wang *et al.* 2012. * Thermo Fisher Scientific; ** New England BioLabs inc.

Table S3. Results of the best fit model selection for each of the RE dataset.

	NLS model equation	Residual standard error	Parameters and P-values	Iterations to convergence	Correlation y vs. predicted values	AIC
Polymorphic loci with up to 2 SNPs shared by at least 90% of samples						
AlfI	Logarithmic $y \sim a + b\ln(x)$	2.49, df=2	a = 38.5, p < 0.001 b = 27.4, p < 0.001	1	0.99	19.8
BcgI	Geometric $y \sim ax^{bx}$	11.08, df=2	a = 47.7, p < 0.01 b = 0.20, p < 0.001	13	0.99	31.8
CspCI	Exponential $y \sim ae(bx)$	16.82, df=2	a = 6.9, p < 0.01 b = 1.1, p < 0.001	9	0.99	35.2
Polymorphic loci with up to 2 SNPs shared by at least 80% of samples						
AlfI	Geometric $y \sim ax^{bx}$	7.87, df = 2	a = 54.4, p < 0.001 b = 0.4, p < 0.001	6	0.99	29.1
BcgI	Geometric $y \sim ax^{bx}$	6.39, df = 2	a = 64.8, p < 0.001 b = 0.26, p < 0.0001	13	0.99	27.4
CspCI	Power-law $y \sim ax^{b}$	91.3, df = 2	a = 16.7, p < 0.1 b = 3.2, p < 0.001	6	0.99	48.7

S1 Code. Nonlinear least squares (NLS) analysis. Available on request