



OPEN

The first high-density genetic map of common cockle (*Cerastoderma edule*) reveals a major QTL controlling shell color variation

Miguel Hermida¹, Diego Robledo², Seila Díaz^{3,4}, Damián Costas⁵, Alicia L. Bruzos^{3,6}, Andrés Blanco¹, Belén G. Pardo¹ & Paulino Martínez¹✉

Shell color shows broad variation within mollusc species and despite information on the genetic pathways involved in shell construction and color has recently increased, more studies are needed to understand its genetic architecture. The common cockle (*Cerastoderma edule*) is a valuable species from ecological and commercial perspectives which shows important variation in shell color across Northeast Atlantic. In this study, we constructed a high-density genetic map, as a tool for screening common cockle genome, which was applied to ascertain the genetic basis of color variation in the species. The consensus genetic map comprised 19 linkage groups (LGs) in accordance with the cockle karyotype ($2n = 38$) and spanned 1073 cM, including 730 markers per LG and an inter-marker distance of 0.13 cM. Five full-sib families showing segregation for several color-associated traits were used for a genome-wide association study and a major QTL on chromosome 13 associated to different color-traits was detected. Mining on this genomic region revealed several candidate genes related to shell construction and color. A genomic region previously reported associated with divergent selection in cockle distribution overlapped with this QTL suggesting its putative role on adaptation.

The common cockle, *Cerastoderma edule*, is a bivalve mollusc naturally distributed along the Northeast Atlantic coast, from Senegal in the South to Norway and Iceland in the North, inhabiting on intertidal soft sediment regions¹. This species has an important ecological role on marine sediment renewal and represents a food source for birds, crustaceans and fish, thus playing an important role for coastal ecosystems and marine communities². The species is considered a delicacy and it is commercially fished mainly in Ireland, United Kingdom, Netherlands, France, Spain and Portugal, where it represents a valuable species to coastal fisheries³.

As bivalve molluscs, the shell is a fundamental part of the cockle, serving as protection against predators, desiccation at intertidal zones or mechanical damage, enabling behaviours such as being swept by currents or burrowing. This multi-layered exoskeleton is constituted mainly of calcium carbonate deposited into an organic matrix of proteins and pigments secreted by specialized epithelial cells on the dorsal mantle⁴. Although a conserved set of regulatory genes appears to underlie mantle progenitor cell specification, the genes that contribute to the formation of the mature shell are diverse⁵. Technical innovations have allowed to discover surprising patterns of shell pigmentation and rapid divergences in the mix of pigments used to achieve similar color patterns^{6,7}. Indeed, the shells of different bivalves are remarkably diverse, and shell pigmentation varies dramatically even within species. Variation in shell color and its pattern may be associated with different biotic or abiotic factors such as predation, substrate, diet or environmental conditions^{8,9}. Since color variation has been reported to be controlled to some extent by genetic factors, shell color might be important for adaptation of bivalve population to selective pressures^{10,11}. Moreover, as a commercialized food resource, shell color can be important for

¹Department of Zoology, Genetics and Physical Anthropology, Acuigen Group, Faculty of Veterinary, Universidade de Santiago de Compostela, Campus of Lugo, 27002 Lugo, Spain. ²The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, UK. ³Genomes and Disease Group, Department of Zoology, Genetics and Physical Anthropology, Center for Research in Molecular Medicine and Chronic Diseases (CiMUS), Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain. ⁴ECOMARE, CESAM-Centre for Environmental and Marine Studies, Department of Biology, University of Aveiro, Santiago University Campus, 3810-193 Aveiro, Portugal. ⁵Centro de Investigación Mariña, Universidade de Vigo, ECIMAT, 36331 Vigo, Spain. ⁶Mosaicism and Precision Medicine Group, Department of Genetics and Genomic Medicine, The Francis Crick Institute, University College of London, London, UK. ✉email: paulino.martinez@usc.es

consumer pleasantness and acceptability, affecting to a certain extent the sale value. Whether common cockle populations show variation in their coloration patterns, if it has an underlying genetic basis and to what extent it could be related to adaptive variation is unknown.

The rapid expansion of next-generation sequencing (NGS) in the last decade has allowed the development of genotyping-by-sequencing (GBS) methods, which have been employed to discover and genotype thousands of single nucleotide polymorphisms (SNPs) in a cost-effective manner, enabling population-scale genetic studies in non-model species¹². These methods, including Restriction-site Associated DNA (RAD) sequencing¹³ and its derivations, ddRAD¹⁴, 2b-RAD¹⁵ or SLAF¹⁶, have been successfully used for high-throughput genotyping in many aquaculture species¹⁷, including several important commercial molluscs¹⁸. GBS methods have been recently applied to understand adaptive variation of common cockle from Northeast Atlantic, and consistent signals of adaptive variation were detected both at microgeographic (dd-RAD;¹⁹) and macrogeographic (2b-RAD;²⁰) scales. GBS have facilitated the construction of high-resolution linkage maps^{21–23}, which are important tools for genome scaffolding and assembly²⁴ and have aided to disentangle the genetic basis of relevant evolutionary or productive traits through quantitative trait locus (QTL) screening^{25,26}. Genetic maps have been used to study the genetic architecture of traits of interest in various bivalve species, such as growth in Zhikong scallop (*Azumapecten farreri*;²⁷), bay scallop (*Argopecten irradians*;²⁸) or Pacific oyster (*Crassostrea gigas*;²⁹), various pearl-quality traits in triangle sail mussel (*Hyriopsis cumingii*;³⁰) and resistance to pathologies in European flat oyster (*Ostrea edulis*;³¹). Previous studies have also identified QTL for shell coloration in several bivalves including Manila clam (*Ruditapes philippinarum*;^{32–34}), hard clam (*Mercenaria mercenaria*;³⁵), Pacific oyster (*Crassostrea gigas*;^{36–39}), black-lip pearl oyster (*Pinctada margaritifera*;⁴⁰), Akoya pearl oyster (*Pinctada fucata*;⁴¹) and Yesso scallop (*Mizuhopecten yessoensis*;^{42,43}), and therefore, similar strategies might be employed to ascertain the genetic component underlying differences in shell coloration in common cockle.

In this study, we investigated the variation of shell color in Northeast Atlantic populations of European common cockle and studied its genetic architecture through a genome-wide association study (GWAS) on several full-sibs families using the first common cockle high-density linkage map here constructed using 2b-RAD SNP genotyping. A major QTL underlying shell color and its pattern in this species was detected on chromosome 13 and several candidate genes identified. This information should be considered as a potential source for adaptive variation in common cockle and could be exploited in breeding programs to adapt production to consumer demands.

Results

Color and pattern variation in common cockle European populations. Shell color showed great diversity among the 270 cockle individuals analysed in Northeast Atlantic populations (Fig. 1). Although a predominant color was displayed in each population, important variation was also observed within the nine cockle beds analysed (Fig. 2).

Yellow (1) was the only color detected in the population from France and was the most abundant in the populations from Denmark, Portugal (Algarve) and UK (Wales); white (2) was predominant in populations from Spain, Netherlands and Germany; brown (4) was the least frequent color in this natural survey, only detected in one of the populations from Portugal (Algarve); and orange (6) was rather common in the other population from Portugal (Aveiro), Netherlands and UK (Plymouth) (Fig. 2). Finally, gray (3) and black (5) colors were only detected in the families studied at hatchery using breeders from NW Spain (see below), so apparently rare in European populations.

Shell patterns in wild adults (Fig. 1) were not as marked as in juvenile samples obtained from crosses in the hatchery (see below) and consequently were not systematically recorded. Nonetheless, different color patterns similar to those observed in the hatchery were also identified: (i) a lighter coloration in the circle-shaped umbo in the population from Portugal (Algarve); (ii) a darker band (black or orange) with blurred boundaries in the lateral area of the shell on the opposite side of the ligament in populations from UK (Wales, black) and Portugal (Aveiro, orange); (iii) changes in the arrangement of the periostracum (protein layer) detected in the ventral margin associated with the last growth rings in all populations, excluding Denmark, where remnants of the periostracum were detected on the entire surface of the shell. Additionally, malformation of the shell that affected the last growth rings, likely related to environmental factors, were detected in individuals from France and Portugal (Algarve).

2b-RAD sequencing. Cockle breeders to produce families were collected in the natural bed of Noia (NW Spain) and transferred to the ECIMAT-CIM-UVigo (Vigo, NW Spain), where five families were used for GWAS on shell color and its pattern; further, the two more numerous families (F6 and F8) were used for genetic map construction. A total of 275 samples were sequenced in three 2b-RAD libraries: (i) the two parents and 97 offspring of F6; (ii) the two parents and 99 offspring of F8; and (iii) 25 offspring from each of F2, F3 and F7. Around ~ 575 million raw reads were obtained in the first two libraries for the two large families: on average ~ 6.9 million for parents (range: 5,635,542–8,343,112) and ~ 2.8 million for offspring (range: 3,680–5,991,704). After filtering, ~ 75% of the reads were retained and aligned to the cockle genome (Tubío et al., unpublished). An important number of reads were discarded due to mapping to two or more genomic positions (40.74%), resulting in ~ 200 million single-site aligned reads: ~ 3 million from each parent (range: 2,101,828 to 3,843,549) and ~ 900,000 from each offspring (range: 1,317 to 2,158,061).

The third run, with the remaining 75 samples, yielded ~ 275 million raw sequences (~ 3.6 million per offspring), and after the filtering and alignment steps, ~ 92 million high-quality aligned reads were retained (~ 1.2 million per offspring).



Figure 1. Representative individuals for each color and pattern identified in shells of *C. edule* from Northeast Atlantic. From up to down and left to right: color phenotypes yellow (1), white (2), gray (3), brown (4), black (5), orange (6), and shell patterns circle (7), line (8) and stripe (9).

Genotyping and linkage map construction. The *gstacks* module using the *marukilow* model applied to all families yielded 318,755 loci, which resulted in 85,078 polymorphic SNPs using the *populations* module. One individual from each of the two families used for mapping showed a low number of valid genotypes (<30% of SNPs genotyped) and they were removed. After quality control, 7,094 and 8,439 SNPs were retained for F6 and F8, respectively. The largest number of filtered SNPs in our study was due to missing genotypes and deviations from Mendelian segregation, which can be explained by the presence of null alleles related to polymorphism in the restriction enzyme targets, as previously reported in molluscs^{44,45}. There were 1,329 common informative SNPs between the two families, which means that in total 14,204 SNPs were used for the construction of the genetic map.

Separate male and female maps were built for each family (Supplemental Tables S1–S4; Supplemental Fig. 1). To achieve an appropriate number of LGs close to the number of cockle chromosomes ($n = 19$), different LOD scores were explored for each genetic map (ranging between 7.0 and 9.0), resulting in 21 LGs in all maps. For the maternal map of F6, 3,514 markers were mapped for a total length of 12,572 cM, whereas the paternal map included 3952 markers spanning 16,692 cM (Supplemental Table S5). In F8, 4,698 and 4,368 markers were mapped in the maternal and paternal maps, spanning 25,017 and 23,266 cM, respectively. Shared markers among parental maps were used to build a single consensus map. As a result, 13,874 SNPs were assigned to 19 LGs in the final common cockle genetic map with a total length of 51,778 cM, in accordance with the 19 cockle chromosomes (Supplemental Tables S5–S6). The length of the maps exceeded that expected based on the genome size considering a standard relationship between physical (Mb) and genetic distance (cM) of 1.2 and a C-value of 1.37 pg⁴⁶. The observed elongation of the genetic maps is the consequence of the high number of markers, several mapping families and the limitations of the software used (JoinMap for mapping and MergeMaps for *consensing*), consistent with previous observations^{21,47,48}.

To build a reliable framework genetic map we used the Regression Mapping approach, a similar approximation to that followed in *C. gigas* by Hedgecock et al.⁴⁹. Accordingly, a total of 831 and 340 markers without

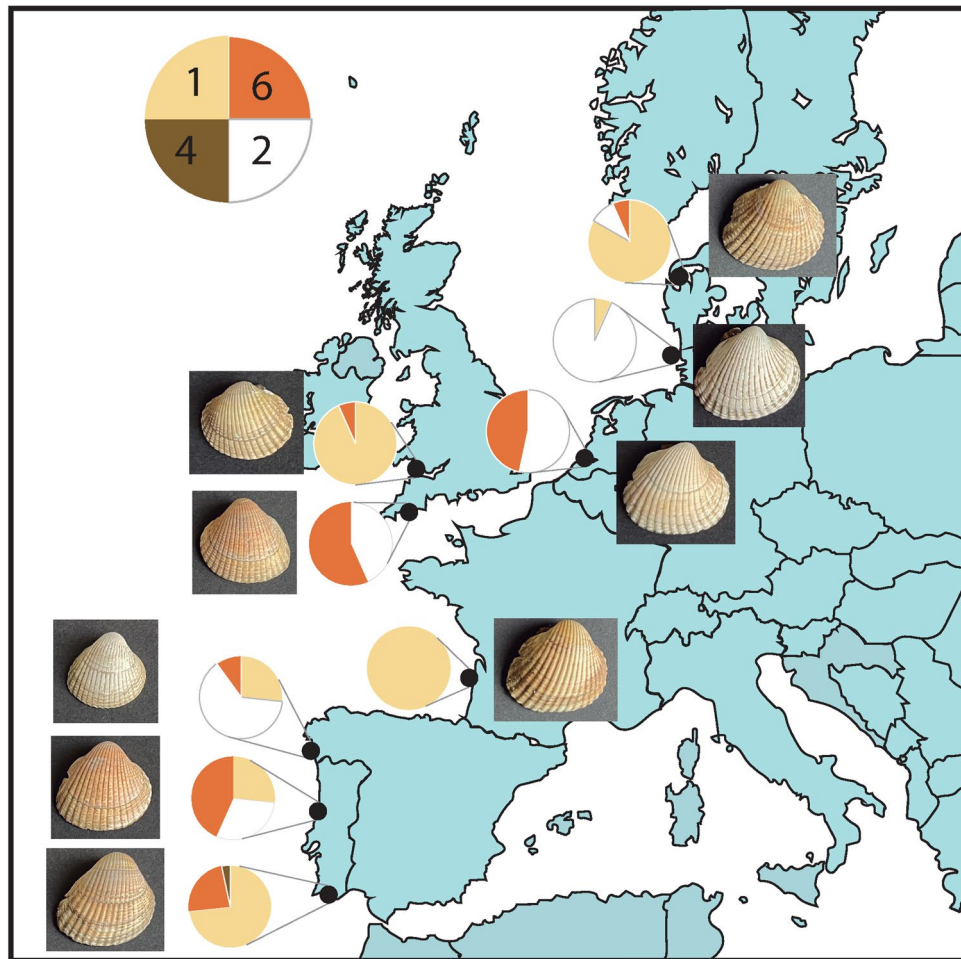


Figure 2. Geographical representation of the European samples of *C. edule*, including pie chart color distributions (yellow (1), white (2), brown (4), orange (6)) and a photograph of a representative individual. From north to south: Nykobing Mors (Denmark), Sylt (Germany), Slikken van Viane (Netherlands), Wales and Plymouth (United Kingdom), Arcachon (France), Baiona (Spain), Aveiro and Algarve (Portugal).

missing data were selected in offspring of F6 and F8, respectively. Separate maps were built for each parent in each family with a LOD score ≥ 5.0 . Graphical representation was not implemented, but the individual groups were merged directly with the “Combine Groups” option of the JoinMap to build a consensus map. The regression of common markers distance was used to correct the distances in the original consensus map to build a new corrected-length consensus map, which included the 19 LGs but with a new total length of 1073 cM (Fig. 3). In the reduced consensus map, the estimated inter-marker distance decreased dramatically (from 4.34 to 0.13 cM), comparable to other genetic linkage maps constructed using 2b-RAD^{50,51}, and the average ratio between physical and genetic distance (0.74 Mb/cM) was also much closer to that expected.

GWAS on shell color type and pattern at hatchery. Shell color and its pattern showed a remarkable variation within and among the five families reared in the same environmental conditions (Fig. 4). Shell color of the offspring of the five families was classified as outlined before (from 1 to 6) and considered as a continuous trait for analyses (Table 1). Five of the six colors identified in common cockle (Fig. 1) were identified in families. Black was the most frequent color (31.6%), but it was detected only in two families (F2 and F8), whereas gray was the least frequent and detected only in F8. White was the only present in all families and the second most abundant in the whole sample (29.8%), and brown was missed only in one family. Orange was not detected in any of the families, only in the wild, as outlined before. Color patterns were only detected in three families and were quite heterogeneous; for instance, stripe was only observed in F6 at a ratio close to 1:1 (Fig. 4; Table 1).

The estimated heritabilities were high for color and stripe, 0.755 and 0.657, respectively, and moderate-high for circle and line, 0.537 and 0.506, respectively. A genetic correlation of almost 1 was observed between color and circle (0.998), while these patterns showed a moderate negative genetic correlation with stripe (-0.327 and -0.379 , respectively) (Table 2). Line did not show significant genetic correlation with any of the other traits. The circle in the umbo was associated with darker shell colors, which suggests a particular pattern not visible in the whitish phenotypes.

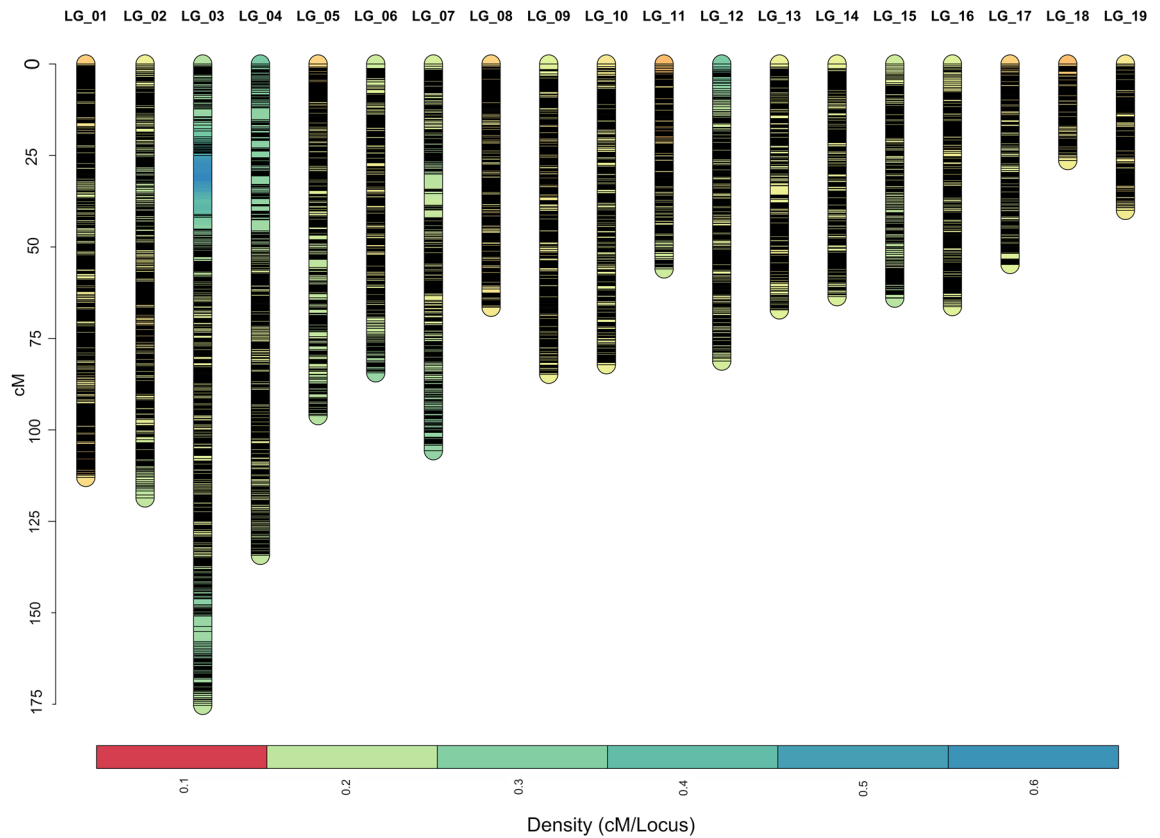


Figure 3. Graphical representation of the consensus map of *C. edule* including 13,874 markers encompassing 1073 cM. The rule in the left indicates length in centimorgans (cM).

The complete dataset contained color phenotypes for 275 individuals from five families, genotyped for 13,874 SNPs mapped in the Consensus Map (between 4,643 and 8,439 informative markers per family). GWAS revealed a highly significant genome-wide QTL for color and stripe, and to a less extent for circle, at chromosome C13 (Fig. 5). Other SNP associations at chromosome-level were detected for circle at C8, for line at C2, C9, C12 and C14, and for stripe at C1, C5, C9, and C17 (Supplemental Table S7).

Gene mining. GWAS identified a convincing QTL associated with color, stripe and circle located at C13 between 14,367,847 and 33,654,270 bp (Fig. 5). The highest significant SNP associated with the three traits was located at 30,286,849 bp, but across this wide region, there were several stretches defined by highly significant SNPs associated with color and stripe, the most significant-associated traits (Fig. 6). The two end subregions were mainly associated with color, and mining around the most significant SNPs (14,778,145 and 33,225,111 bp \pm 500 kb) revealed 16 and 17 annotated genes, respectively (Supplemental Table S7). Half of the genes in the first window, related to shell architecture and color, clustered on a \sim 250 kb region (Fig. 6; Supplemental Table S7): five were related to chitin binding (three microfibril-associated glycoprotein 4, one including a fibrinogen domain, and one DNA damage-regulated autophagy modulator), one to calcium binding (ependymin-related), one to mucin secretion, and one to ammonium transport. The window at the other end (33,225,111 kb) included two genes related to iron binding (two steroid 17- α -hydroxylase/17,20 lyase-like) and another one to calcium transport (phosphatidylinositol 4,5-bisphosphate phosphodiesterase).

On the other hand, two consecutive subregions associated with stripe color pattern were located around the most significant SNPs at 22,683,835 and 30,286,849 bp (\pm 500 kb) on the same chromosome and comprised a total of 36 annotated genes. These group of genes included three cell membrane organic transporters (solute carrier organic anion transporter family member 4A1-like, urea-proton symporter DUR3-like and sodium/glucose cotransporter 4-like); one chitin-binding (sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1-like); several involved in transit across the membrane by endo-exocytosis mechanisms (prenylcysteine oxidase 1-like, glucose-6-phosphate exchanger SLC37A4-like, and AP-2 complex subunit alpha-2-like); and one related to iron chelation (putative ferric-chelate reductase 1) (Supplemental Table S7).

Discussion

Molluscs represent a highly diverse Phylum of invertebrates comprising an estimated number of 200,000 species, distributed across almost every type of habitat worldwide¹⁸. They have key roles as ecosystem engineers, water filtering and pollution monitoring, jewelry and, of course, as an important food source. More than 17 million metrics tons of molluscs were farmed worldwide in 2018 and most of this production concentrated in

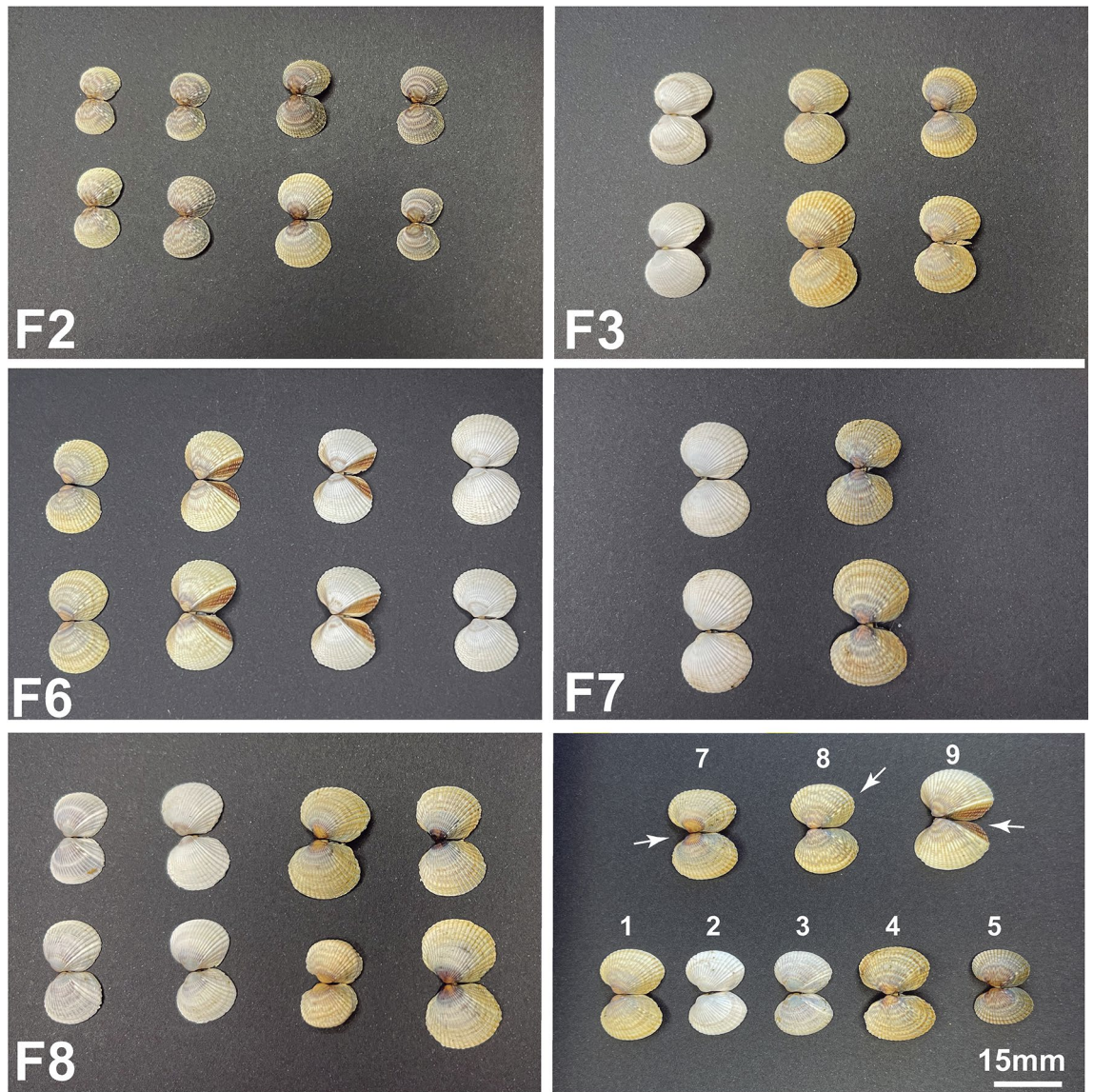


Figure 4. Photograph composition showing representative individuals of the five families (F2, F3, F6, F7 and F8) of *C. edule* used for GWAS on shell color and pattern. Last panel: summary of the variation within and between families: yellow (1), white (2), gray (3), brown (4), black (5), circle (7), line (8) and stripe (9).

Family	Shell color					Shell color pattern		
	Yellow	White	Gray	Brown	Black	Circle	Line	Stripe
Family 2	3	1	0	8	13	7	11	0
Family 3	7	12	0	6	0	0	0	0
Family 6	24	45	0	31	0	0	0	54
Family 7	0	13	0	12	0	0	0	0
Family 8	4	11	11	0	74	36	8	0
Total (%)	38 (13.9)	82 (29.8)	11 (4.0)	57 (20.7)	87 (31.6)	43 (37.1)	19 (16.4)	54 (46.6)

Table 1. Distribution of shell color and pattern phenotypes in the offspring of the five families of *C. edule*.

a handful of species of the class Bivalvia⁵². Bivalve aquaculture mainly relies on extensive farming based on the collection of wild seed and harvesting in natural beds, which means that wild populations are under important human alterations⁴⁵.

Shell constitutes a main structure of mollusc anatomy that protects them against predators and desiccation, but that also plays other important functions depending on species. Shells are secreted by the mantle, so color

Trait	Color	Circle	Line	Stripe
Color		0.465***	0.147*	-0.501***
Circle	0.998		0.049 ^{n.s.}	-0.208***
Line	0.173	0.137		-0.130*
Stripe	-0.379	-0.327	-0.217	

Table 2. Phenotypic and genetic correlations (upper triangle and lower triangle, respectively), between color traits of *C. edule* included in this study. (***) $p < 0.001$, (*) $p < 0.05$.

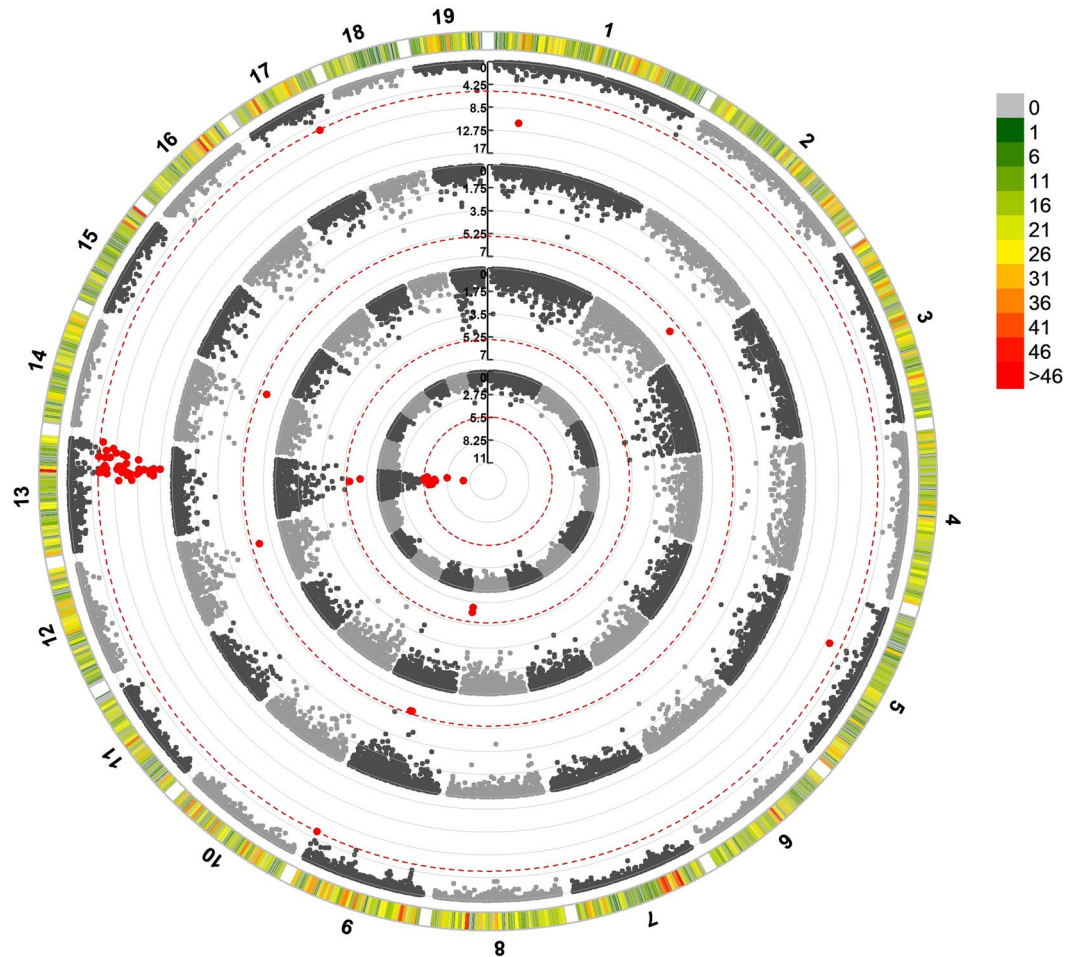


Figure 5. Circular Manhattan plot for SNP significantly associated with shell color and pattern phenotypes in *C. edule* based on the Mixed Linear Model (MLM). SNP-GWAS circle representations: inner-most: color; first-middle: circle; second-middle: line; and outer-most: stripe. For Manhattan plot, Y-axis represents $-\log_{10}(p\text{-value})$ of the association with each SNP and X-axis is physical position in bp. The dashed red lines represent the Bonferroni threshold at genome-wide level ($p \leq 0.05$). SNP markers are represented by gray dots and those above this threshold in red dots.

and its pattern are mainly determined by pigments produced by this tissue, although the microstructure of the shell may also contribute to coloration^{4,53}. Despite recent efforts, there is still a knowledge gap regarding the genetics underpinning shell color in Mollusca and its potential role on adaptation. While it is well-known that shell color can be under genetic control, biotic and abiotic factors, such as diet, temperature, salinity or pH can also play a role, and the interaction between them is not well understood yet⁸. To address the genetic architecture of complex traits, such as shell color, comprehensive genomics approaches are essential.

Genomic resources have increased exponentially in the last decade as consequence of the lowering cost of sequencing technologies and the new bioinformatic tools that enabled the assembly of genomes at chromosome level, the construction of high-density genetic maps and the genotyping of millions of SNPs for genomic screening⁵⁴. Genetic maps are essential for the identification of genomic regions underlying phenotypic variation

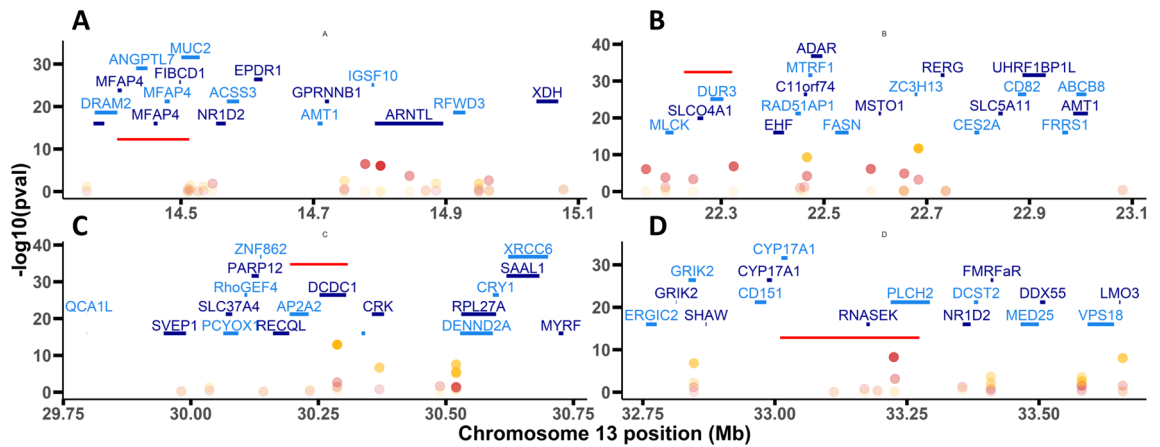


Figure 6. Manhattan plots for SNP significantly associated with color (**A, D**) and stripe pattern (**B, C**) phenotypes in four chromosome 13 subregions of *C. edule* based on the Mixed Linear Model (MLM). Y-axis represents $-\log_{10}(p\text{-value})$ of the association with each SNP and X-axis its physical position in bp. Codes of genes mined in those regions are shown with underlined bars representing their length in Mb; yellow and red dots highlight associated SNPs with stripe and color, respectively; red bars at each plot show the most interesting candidates in the subregion.

for relevant traits under domestic or natural selection^{44,45,55}. The first mollusc genetic maps were published in *Crassostrea virginica*⁵⁶ and *Crassostrea gigas*^{57,58} using AFLPs and microsatellites, respectively, but the lowering cost of sequencing technologies enabled genotyping thousands of SNPs for improving map density^{29,32}. The common cockle genetic map here constructed comprehends 13,874 markers with an inter-marker distance of 0.13 cM and comprises the 19 expected LGs matching with the haploid karyotype of the species⁵⁹, being, to our knowledge, the densest genetic map published to date in molluscs. Besides its importance for genomic screening, the common cockle genetic map is an invaluable tool for genome scaffolding, as has been previously reported in other species^{23,60}, and in fact, five of the seven mollusc genomes assembled at chromosome-level took advantage of high-density linkage maps⁴⁵.

We used the common cockle genetic map to ascertain the genetic component underlying the broad diversity of shell color and pattern on the wild populations of this species in the Northeast Atlantic. Ricardo et al.^{61,62} reported an important variation in shell ion composition in common cockle, apparently related to environmental factors, that allowed tracing back the geographic origin of specimens, but they did not associate this variation with color. Most phenotypes observed in wild populations in our study could be identified in the families produced at hatchery, although their intensity was somewhat faded, likely due to environmental factors operating across the lifespan of the adult specimens collected. Indeed, differentiation of individuals by color in families was very clear, sometimes resembling single gene Mendelian segregation, and heritabilities for all color traits evaluated were high ($h^2 > 0.5$), supporting a significant genetic component underlying color variation in common cockle. Both color, circle and stripe showed highly significant genetic and phenotypic correlations, suggesting that the same genes (or genomic regions) could underlie color variation for these three traits. Moreover, the fact that most phenotypes observed in the Atlantic distribution appear to be segregating in a single population from NW Spain supports an important intrapopulation variation in common cockle and suggest that more detailed studies across the full lifespan could disclose more variation than observed in our preliminary screening.

In accordance with these observations, the GWAS performed on five full-sib families produced at hatchery identified a major QTL at C13 for two of the four traits evaluated (color and stripe). This region encompassed ~13 Mb, although with different stretches for the same or the different traits studied, which suggests the existence of a broad gene cluster related to color pigmentation in common cockle with different genes playing diverse functions on similar traits. A total of 69 annotated genes were found on this region after mining the cockle genome (Tubío et al., unpublished). We identified a notable proportion of genes related to ion binding and transport/secretion across the cell membrane, such as ammonium or organic transporters, calcium binding or iron chelation, mucin production and several related to endo-exocytosis mechanisms, which play an important role in the development of the shell and that has been related with shell color in other molluscs^{35,38,41,42,63,64}. Moreover, shells are secreted by the mantle in a process called biomineralization, where chitin represents an important component^{40,65,66}. We could identify six chitin-binding related genes, four of them in a small genomic region encompassing ~110 kb including three microfibril-associated glycoprotein 4 genes. Genes related to chitin and calcium metabolism involved in different shell color lines have been reported in *Mizuhopecten yessoensis*⁴². Nevertheless, it is important to note that previous studies have shown that some of the genes related to shell architecture and color are species-specific^{8,53}, so further studies should be conducted in the future to ascertain the roles of the genes located at the C13 cluster for a deep understanding of color type and pattern diversity in common cockle.

A reflection on the putative adaptive role of color diversity in the common cockle is worth a final thought. While shell color can be important for the adaptation of bivalve populations to selective pressures, this might not be a major factor in molluscs that live buried in the sediment, such as *C. edule*. Nevertheless, cockles live in the intertidal zone and constitute the feed for different predators, such as birds, mammals and crustaceans, among

others. Moreover, a broad color diversity exists in this species across its full distribution range underpinned by a substantial genetic variation, as the high heritabilities estimated in our study demonstrate, so its putative adaptive role should deserve further studies. Interestingly, the C13 color associated QTL overlaps with a genomic region in the same chromosome which showed very consistent signals of divergent selection in the whole Northeast Atlantic and in the Northern Region (above the Ushant Front), including several outliers above the neutral background and highly significant linkage disequilibrium suggestive of selective sweeping²⁰.

Conclusion

Here we presented a high-density genetic map in common cockle, the first reported in the species and, to our knowledge, the highest dense map reported in molluscs to date. The consistency of the map was shown by fitting the number of linkage groups to the haploid chromosome number of the species and by the consistent result of the GWAS on shell color and pattern. This map was used to ascertain the genetic component and architecture underlying the broad color diversity observed in common cockle in the Northeast Atlantic. High heritabilities were estimated for all traits evaluated, supporting an important genetic component underlying coloration pattern. This genetic variation was mainly associated with a single genomic region at C13, where a cluster of genes related to specific enriched functions on shell architecture and color were detected. Our preliminary results in the wild suggest a potential adaptive role of the color variation observed and highlights the importance of deeper studies at population level across the cockle lifespan to understand the significance of the variation observed.

Material and methods

Cockle families. In May 2018, 300 mature adult *C. edule* cockles were collected in Noia, Galicia (NW Spain) and transferred to the ECIMAT-CIM-UVigo marine facilities (Vigo, NW Spain). Cockles were kept individually in glasses with 0.3L of 1 μ m filtered seawater at 20 °C. Spawning was induced by thermal shocks between 10 and 22 °C for 10 h and the quality of the oocytes and sperm was evaluated under a light microscope. Controlled fertilization was carried out by adding sperm to oocytes, one male x one female, at a ratio of 1:10. D-shaped larvae were obtained 24 h after fertilization with a transformation rate from trochophore larvae of 42% \pm 19. Following this protocol, a total of eleven full-sib families were obtained by crossing one male x one female, involving a total of 11 females and 7 males.

Larvae from each family were cultured in individual 150 L cylindrical-conical tanks at a density of 8 \pm 3 larvae mL⁻¹ with sea water filtered at 1 μ m and treated with UV, slight aeration, and temperature 19.0 \pm 1.4 °C in an open circuit with a renewal of 5% volume / hour. The diet consisted of *Tisochrysis lutea* (ECC038), *Chaetoceros neogracile* (ECC007), *Phaeodactylum tricoratum* (ECC028) and *Rhodomonas lens* (ECC030) in a ratio of 1:1:1:1 (according to the cell count), and *Tetraselmis suecica* (ECC036) was included from the seventh day of culture. The daily diet was administered automatically every 4 h in 6 daily intakes, maintaining a constant concentration in the tank of 20–40 cells μ L⁻¹. At 14 days post-fertilization (dpf), pediveliger larvae from each family were transferred to separate 50 L tanks in suspended baskets with constant aeration, temperature 18.4 °C \pm 0.5 and a renewal rate of 50 L day⁻¹. The animals were fed with the same diet as described above but maintaining a constant density of 168 \pm 48 seeds cm². Metamorphosis of larvae took place in those tanks.

At 112 dpf one hundred individuals from each of two families, F6 (12.74 \pm 0.73 mm) and F8 (12.23 \pm 1.48 mm), were selected for genetic mapping and for GWAS on color patterns, while twenty-five individuals from each of three additional families, F2 (9.82 \pm 0.85 mm), F3 (12.58 \pm 0.90 mm) and F7 (13.12 \pm 1.67 mm), were sampled for increasing statistical power of GWAS. The meat of each individual was fixed in pure ethanol and sent to the Genomics Platform of University of Santiago de Compostela (Campus de Lugo) for DNA extraction.

Color variation of common cockles. After the visual inspection of all the shells in the study, shells were classified according with their color into six phenotypic classes: yellow (1), white (2), grey (3), brown (4), black (5), orange (6). Further, specific color patterns of the shell were consistently identified and recorded as three other traits: (i) a circle in the umbo, differentiated from the rest of the shell and generally yellow (circle); (ii) a broken white line (line); and (iii) a stripe with white line edge (stripe) (Fig. 1). A typical individual from each class was selected to define a standard pattern to categorize the phenotype of each shell in the study. This evaluation was performed by a single observer, who revised and scored all the shells in two independent rounds. Accordingly, color and pattern of the shell were assigned to all hatchery specimens (F2, F3, F6, F7, F8 families), while mainly only the color was assessed in the European wild population samples provided by the Scuba Cancers Project (ERC-2016-STG). Despite similar color patterns were detected in the hatchery and wild individuals, its classification was not straightforward in wild specimens, likely due to environmental factors across their life span, and therefore they were not systematically recorded.

2b-RAD library construction and sequencing. DNA was extracted from the whole meat using the E.Z.N.A. E-96 mollusc DNA kit (Omega Bio-tek) following manufacturer recommendations. Library preparation followed the 2b-RAD protocol¹⁵ with slight modifications²¹. Briefly, DNA samples were adjusted to 80 ng μ L⁻¹ and digested using the IIB type restriction enzyme *AlfI* (Thermo Fisher). As a result, the genome was cut in fragments of 36 bp of length, with the restriction enzyme recognition site in the middle. Specific adaptors, also including individual sample barcodes, were ligated and the resulting fragment amplified. After PCR purification, samples were quantified using Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA, USA) and equimolarly pooled. The pools were sequenced on a NextSeq 500 Illumina sequencer using the 50 bp single-end chemistry in the facilities of FISABIO Sequencing and Bioinformatics Service (Valencia, Spain). The two bigger families including ~ 100 offspring per family, with parents at double concentration, were multiplexed each in one run, whereas the other seventy-five samples, without parents, were multiplexed in a third run.

Data filtering and genotyping. Raw reads were first demultiplexed according to the individual barcodes ligated during library construction. Then, reads were filtered in three consecutive steps: (i) trimmed to 36 nucleotides (length of *AlfI* generated fragments) and discarding reads below this length; (ii) removing reads without the *AlfI* recognition site in the correct position; and (iii) removing reads with uncalled nucleotides or a mean quality score below 20 in a sliding window of 9 nucleotides. Custom Perl scripts (available on demand) were used in the first two steps, and the *process_radtags* module in STACKS v2.0⁶⁷ was used for the latter (-c -q -w 0.25 -s 20). Bowtie 1.1.2⁶⁸ was used to align the filtered reads against the recently assembled reference genome of the species (Tubío et al., unpublished), allowing three mismatches and a unique alignment (-v 3 -m 1 -sam), so reads aligning to two or more sites were discarded. The *sam* output files were converted to *bam* files and appropriately sorted to feed the *gstacks* module in STACKS, using the *marukilow* model to call variants and genotypes. The collection of putative SNPs was exported using the *populations* module.

Linkage map construction. Files containing the putative SNPs of the two families selected for genetic mapping (F6 and F8) were processed to conform an appropriate dataset for mapping. Those SNPs not informative in the parents of each family or with missing data in one or both parents were filtered, as well as those SNPs with extreme deviations from Mendelian segregation ($p < 0.001$) or genotyped in less than 60% of the offspring.

Genotypes of retained SNPs were properly coded as Cross Pollinator (CP) cross type with unknown linkage phase to build genetic maps using JoinMap 4.1 genetic mapping software⁶⁹. Firstly, markers were associated to their linkage groups using the Grouping function of JoinMap based on a series of LOD scores increasing by 1, from 4.0 to 9.0. The LOD score was then selected in each family based on the number of chromosomes of the common cockle karyotype⁵⁹. Markers in linkage groups with less than 10 markers and unlinked markers were excluded for further analysis. Secondly, marker ordering was performed using the Maximum Likelihood (ML) algorithm implemented in JoinMap. Default parameters were used except the chain length that was increased from 10,000 to 20,000, and multipoint estimation of recombination frequency were in general increased to ensure convergence: length of burn-in chain, 20,000; number of Monte Carlo EM cycles, 10; and chain length per EM cycles, 5000. The *Kosambi* mapping function was used to compute centiMorgans (cM) map distances in individual female and male maps in each family. Finally, female, and male consensus maps (from both families), and a species consensus map were built using MergeMap⁷⁰.

The combination of the ML mapping algorithm implemented in JoinMap with missing data and genotyping errors, as well as the use of MergeMap to merge the parental maps and the large number of markers, can derive in inflated map lengths (see e.g.^{71–73}). Nevertheless, ML is more powerful and robust in ordering markers in CP populations compared to Regression Mapping (RM)⁷⁴. To obtain the most accurate linkage maps, a mixed strategy was implemented: (i) a framework genetic map was constructed using RM only with markers without missing data; (ii) then, a linear regression between ML and RM distances using the same marker pairs was performed; and (iii) this value was used to adjust the map distances in the ML ordered consensus maps. All maps were drawn using MapChart ver. 2.3⁷⁵ and LinkageMapView⁷⁶.

GWAS analysis. For the study of the genetic parameters underlying color variation, the information of five families was used to estimate genetic correlations and heritabilities using hibelup v1.3.1 (<https://github.com/xiaol-ei-lab/hibelup>). The same set of families was used for a genome wide association study to ascertain the association between phenotypic traits and SNPs across the cockle's genome. The genotypes of mapped markers in the Consensus Map and the phenotypes of offspring were used to look for association in the two families used for mapping (F6 and F8), but in addition, to increase statistical power, the same markers were used in three additional families (F2, F3 and F7). In this case the genotypes were also filtered by minimum allele frequency (-maf 0.05) and minimum number of genotypes (-geno 0.5) using PLINK 1.9 (www.cog-genomics.org/plink/1.9;⁷⁷).

A Mixed Linear Model⁷⁸ was applied to the complete dataset of filtered genotypes and the corresponding color phenotypes catalogued as mentioned above using rMVP⁷⁹, a parallel accelerated tool for GWAS implemented in R⁸⁰. The three color pattern traits (circle, line and stripe) were coded as binomial traits (presence /absence). The kinship matrix was previously computed following VanRaden⁸¹, and the EMMA method was employed to the variance components analysis⁸². rMVP and the *qqman* R package⁸³ were used to plot the results.

Gene mining. Coding genes included in a genomic window defined around the most significant SNPs (± 500 kb) detected in GWAS for each trait were retrieved using the common cockle genome (Tubío et al., unpublished).

Data availability

2b-RAD sequencing data of cockle's families are linked to the SRA project PRJNA862862 and will be released on 2022-11-01, available at the following link <https://www.ncbi.nlm.nih.gov/sra/PRJNA862862> (temporary submission ID: SUB11861213).

Received: 12 May 2022; Accepted: 23 September 2022

Published online: 10 October 2022

References

- Hayward, P. J. & Ryland, J. S. *Handbook of the Marine Fauna of North-West Europe* (Oxford University Press, 1995).
- Norris, K., Bannister, R. C. A. & Walker, P. W. Changes in the number of oystercatchers *Haematopus ostralegus* wintering in the Burry Inlet in relation to the biomass of cockles *Cerastoderma edule* and its commercial exploitation. *J. Appl. Ecol.* **35**, 75–85 (1998).

3. Mahony, K. E. *et al.* Mobilisation of data to stakeholder communities. Bridging the research-practice gap using a commercial shellfish species model. *PLoS ONE* **15**(9), e0238446. <https://doi.org/10.1371/journal.pone.0238446> (2020).
4. Clark, M. S. *et al.* Deciphering mollusc shell production: the roles of genetic mechanisms through to ecology, aquaculture and biomimetics. *Biol. Rev.* **95**, 1812–1837 (2020).
5. McDougall, C. & Degnan, B. M. The evolution of mollusc shells. *Dev. Biol.* **7**, e313. <https://doi.org/10.1002/wdev.313> (2018).
6. Miyamoto, H. *et al.* The diversity of shell matrix proteins: genome-wide investigation of the pearl oyster, *Pinctada fucata*. *Zool. Sci.* **30**, 801–816 (2013).
7. Williams, S. T. *et al.* Colorful seashells: Identification of haem pathway genes associated with the synthesis of porphyrin shell color in marine snails. *Ecol. Evol.* **7**, 10379–10397 (2017).
8. Williams, S. T. Molluscan shell color. *Biol. Rev.* **92**, 1039–1058 (2017).
9. Vendrami, D. L. J. *et al.* RAD sequencing sheds new light on the genetic structure and local adaptation of European scallops and resolves their demographic histories. *Sci. Rep.* **9**, 7455 (2019).
10. Vendrami, D. L. J. *et al.* RAD sequencing resolves fine-scale population structure in a benthic invertebrate: Implications for understanding phenotypic plasticity. *R. Soc. Open Sci.* <https://doi.org/10.1098/rsos.160548> (2017).
11. Ding, J. *et al.* Identification of shell-color-related microRNAs in the Manila clam *Ruditapes philippinarum* using high-throughput sequencing of small RNA transcriptomes. *Sci. Rep.* **11**, 8044. <https://doi.org/10.1038/s41598-021-86727-9> (2021).
12. Davey, J. W. *et al.* Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* **12**, 499–510 (2011).
13. Baird, N. A. *et al.* Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**, e3376. <https://doi.org/10.1371/journal.pone.0003376> (2008).
14. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135. <https://doi.org/10.1371/journal.pone.0037135> (2012).
15. Wang, S., Meyer, E., McKay, J. K. & Matz, M. V. 2b-RAD: A simple and flexible method for genome-wide genotyping. *Nat. Methods* **9**, 808–810 (2012).
16. Sun, X. *et al.* SLAF-seq: An efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE* **8**, e58700. <https://doi.org/10.1371/journal.pone.0058700> (2013).
17. Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P. & Houston, R. Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev. Aquac.* **10**, 670–682 (2018).
18. Gomes-dos-Santos, A., Lopes-Lima, M., Castro, L. F. C. & Froufe, E. Molluscan genomics: the road so far and the way forward. *Hydrobiologia* **847**, 1705–1726 (2020).
19. Coscia, I. *et al.* Fine-scale seascape genomics of an exploited marine species, the common cockle *Cerastoderma edule*, using a multimodelling approach. *Evol. Appl.* **13**, 1854–1867 (2020).
20. Vera, M. *et al.* Genomic survey of edible cockle (*Cerastoderma edule*) in the Northeast Atlantic: A baseline for sustainable management of its wild resources. *Evol. Appl.* **15**, 262–285 (2022).
21. Maroso, F. *et al.* High-density linkage maps from 31 full-sibling families of turbot (*Scophthalmus maximus*) provide insights into recombination patterns and chromosome rearrangements throughout a newly refined genome assembly. *DNA Res.* **25**, 439–450 (2018).
22. Dong, C. *et al.* High-density linkage map and mapping for sex and growth-related traits of largemouth bass (*Micropterus salmoides*). *Front. Genet.* **10**, 960. <https://doi.org/10.3389/fgene.2019.00960> (2019).
23. de la Herrán, R. *et al.* A chromosome-level genome assembly enables the identification of the follicle stimulating hormone receptor as the master sex determining gene in *Solea senegalensis*. Preprint at <https://www.biorxiv.org/content/https://doi.org/10.1101/2022.03.02.482245v1> (2022).
24. Fierst, J. L. Using linkage maps to correct and scaffold de novo genome assemblies: Methods, challenges, and computational tools. *Front. Genet.* **6**, 220. <https://doi.org/10.3389/fgene.2015.00220> (2015).
25. Aslam, M. L. *et al.* Genetic variation, GWAS and accuracy of prediction for host resistance to *Sparicotyle chrysophrii* in farmed gilthead sea bream (*Sparus aurata*). *Front. Genet.* **11**, 594770. <https://doi.org/10.3389/fgene.2020.594770> (2020).
26. Yin, X. & Hedgecock, D. Overt and concealed genetic loads revealed by QTL mapping of genotype-dependent viability in the Pacific oyster *Crassostrea gigas*. *Genetics* <https://doi.org/10.1093/genetics/iyab165> (2021).
27. Zhan, A. *et al.* Construction of microsatellite-based linkage maps and identification of size-related quantitative trait loci for Zhikong scallop (*Chlamys farreri*). *Animal Genet.* **40**, 821–831 (2009).
28. Li, H., Liu, X. & Zhang, G. A consensus microsatellite-based linkage map for the hermaphroditic bay scallop (*Argopecten irradians*) and its application in size-related QTL analysis. *PLoS ONE* **7**, e46926. <https://doi.org/10.1371/journal.pone.0046926> (2012).
29. Li, C. *et al.* Construction of a high-density genetic map and fine QTL mapping for growth and nutritional traits of *Crassostrea gigas*. *BMC Genomics* **19**, 626. <https://doi.org/10.1186/s12864-018-4996-z> (2018).
30. Bai, Z.-Y., Han, X.-K., Liu, X.-J., Li, Q.-Q. & Li, J.-L. Construction of a high-density genetic map and QTL mapping for pearl quality-related traits in *Hyriopsis cumingii*. *Sci. Rep.* **6**, 32608. <https://doi.org/10.1038/srep32608> (2016).
31. Harrang, E. *et al.* Can survival of European flat oysters following experimental infection with *Bonamia ostreae* be predicted using QTLs?. *Aquaculture* **448**, 521–530 (2015).
32. Nie, H. *et al.* Construction of a high-density genetic map and quantitative trait locus mapping in the Manila clam *Ruditapes philippinarum*. *Sci. Rep.* **7**, 229. <https://doi.org/10.1038/s41598-017-00246-0> (2017).
33. Nie, H. *et al.* Transcriptome analysis reveals the pigmentation related genes in four different shell color strains of the Manila clam *Ruditapes philippinarum*. *Genomics* **112**, 2011–2020 (2020).
34. Nie, H. *et al.* Transcriptome analysis reveals the pigmentation-related genes in two shell color strains of the Manila clam *Ruditapes philippinarum*. *Anim. Biotechnol.* **32**, 439–450 (2021).
35. Hu, Z. *et al.* Transcriptome analysis of shell color-related genes in the hard clam *Mercenaria mercenaria*. *Comp. Biochem. Physiol.-D Genom. Proteom.* **31**, 100598. <https://doi.org/10.1016/j.cbd.2019.100598> (2019).
36. Feng, D., Li, Q., Yu, H., Kong, L. & Du, S. Transcriptional profiling of long non-coding RNAs in mantle of *Crassostrea gigas* and their association with shell pigmentation. *Sci. Rep.* **8**, 1436. <https://doi.org/10.1038/s41598-018-19950-6> (2018).
37. Song, J. *et al.* Mapping genetic loci for quantitative traits of golden shell color, mineral element contents, and growth-related traits in Pacific oyster (*Crassostrea gigas*). *Mar. Biotechnol.* **20**, 666–675 (2018).
38. Wang, J. *et al.* An integrated genetic map based on EST-SNPs and QTL analysis of shell color traits in Pacific oyster *Crassostrea gigas*. *Aquaculture* **492**, 226–236 (2018).
39. Han, Z. *et al.* QTL mapping for orange shell color and sex in the Pacific oyster (*Crassostrea gigas*). *Aquaculture* **530**, 735781. <https://doi.org/10.1016/j.aquaculture.2020.735781> (2021).
40. Lemer, S., Saulnier, D., Gueguen, Y. & Planes, S. Identification of genes associated with shell color in the black-lipped pearl oyster *Pinctada margaritifera*. *BMC Genomics* **16**, 568 (2015).
41. Xu, M., Huang, J., Shi, Y., Zhang, H. & He, M. Comparative transcriptomic and proteomic analysis of yellow shell and black shell pearl oysters *Pinctada fucata martensii*. *BMC Genomics* **20**, 469. <https://doi.org/10.1186/s12864-019-5807-x> (2019).
42. Ding, J. *et al.* Transcriptome sequencing and characterization of Japanese scallop *Patinopecten yessoensis* from different shell color lines. *PLoS ONE* **10**, e0116406. <https://doi.org/10.1371/journal.pone.0116406> (2015).

43. Zhao, L. *et al.* A genome-wide association study identifies the genomic region associated with shell color in Yesso Scallop *Patinopecten yessoensis*. *Mar. Biotechnol.* **19**, 301–309 (2017).
44. Saavedra, C. & Bachère, E. Bivalve genomics. *Aquaculture* **256**, 1–14 (2006).
45. Hollenbeck, C. M. & Johnston, I. A. Genomic tools and selective breeding in molluscs. *Front. Genet.* **9**, 253. <https://doi.org/10.3389/fgene.2018.00253> (2018).
46. Gregory, T. R. Animal genome size database. <http://www.genomesize.com> (2022).
47. Cartwright, D. A., Troggio, M., Velasco, R. & Gutin, A. Genetic mapping in the presence of genotyping errors. *Genetics* **176**, 2521–2527 (2007).
48. Mester, D., Ronin, Y., Schnable, P., Aluru, S. & Korol, A. Fast and accurate construction of ultra-dense consensus genetic maps using evolution strategy optimization. *PLoS ONE* **10**, e0122485. <https://doi.org/10.1371/journal.pone.0122485> (2015).
49. Hedgecock, D., Shin, G., Gracey, A. Y., Van Den Berg, D. & Samanta, M. P. Second-generation linkage maps for the Pacific oyster *Crassostrea gigas* reveal errors in assembly of genome scaffolds. *G3 Genes Genom. Genet.* **5**, 2007–2019 (2015).
50. Jiao, W. E. *et al.* High-resolution linkage and quantitative trait locus mapping aided by genome survey sequencing: Building up an integrative genomic framework for a bivalve mollusc. *DNA Res.* **21**, 85–101 (2014).
51. Shi, Y. *et al.* High-density single nucleotide polymorphisms linkage and quantitative trait locus mapping of the pearl oyster *Pinctada fucata martensii* Dunker. *Aquaculture* **434**, 376–384 (2014).
52. FAO. Food and Agriculture Organization of the United Nations. FAO Yearbook of Fishery and Aquaculture Statistics. <https://www.fao.org/fishery/en/statistics/yearbook> (2022).
53. Saenko, S. V. & Schilthuizen, M. Evo-devo of shell color in gastropods and bivalves. *Curr. Opin. Genet. Dev.* **69**, 1–5 (2021).
54. Xu, P., David, L., Martínez, P. & Yue, G. H. Editorial: Genetic dissection of important traits in aquaculture: Genome-scale tools development, trait localization and regulatory mechanism exploration. *Front. Genet.* **11**, 642. <https://doi.org/10.3389/fgene.2020.00642> (2020).
55. Tan, K., Zhang, H. & Zheng, H. Selective breeding of edible bivalves and its implication of global climate change. *Rev. Aquac.* **12**, 2559–2572 (2020).
56. Yu, Z. & Guo, X. Genetic linkage map of the eastern oyster *Crassostrea virginica* Gmelin. *Biol. Bull.* **204**, 327–338 (2003).
57. Hubert, S. & Hedgecock, D. Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. *Genetics* **168**, 351–362 (2004).
58. Li, L. & Guo, X. AFLP-based genetic linkage maps of the Pacific oyster *Crassostrea gigas* Thunberg. *Mar. Biotechnol.* **6**, 26–36 (2004).
59. Insua, A. & Thiriou-Quiévreux, C. The characterization of *Ostrea denselamellosa* (Mollusca, Bivalvia) chromosomes: karyotype, constitutive heterochromatin and nucleolus organizer regions. *Aquaculture* **97**, 317–325 (1991).
60. Martínez, P. *et al.* A genome-wide association study, supported by a new chromosome-level genome assembly, suggests *sox2* as a main driver of the undifferentiated ZZ/ZW sex determination of turbot (*Scophthalmus maximus*). *Genomics* **113**, 1705–1718 (2021).
61. Ricardo, F., Pimentel, T., Génio, L. & Calado, R. Spatio-temporal variability of trace elements fingerprints in cockle (*Cerastoderma edule*) shells and its relevance for tracing geographic origin. *Sci. Rep.* **7**, 3475. <https://doi.org/10.1038/s41598-017-03381-w> (2017).
62. Ricardo, F. *et al.* Assessing the elemental fingerprints of cockle shells (*Cerastoderma edule*) to confirm their geographic origin from regional to international spatial scales. *Sci. Total Environ.* **814**, 152304. <https://doi.org/10.1016/j.scitotenv.2021.152304> (2022).
63. Teng, W., Cong, R., Que, H. & Zhang, G. D. novo transcriptome sequencing reveals candidate genes involved in orange shell coloration of bay scallop *Argopecten irradians*. *J. Oceanol. Limnol.* **36**, 1408–1416 (2018).
64. Liao, Z. *et al.* Microstructure and in-depth proteomic analysis of *Perna viridis* shell. *PLoS ONE* **14**, e0219699. <https://doi.org/10.1371/journal.pone.0219699> (2019).
65. Schönitzer, V. & Weiss, I. M. The structure of mollusc larval shells formed in the presence of the chitin synthase inhibitor Nikomycin Z. *BMC Struct. Biol.* **7**, 71. <https://doi.org/10.1186/1472-6807-7-71> (2007).
66. Furuhashi, T., Schwarzinger, C., Miksik, I., Smr, M. & Beran, A. Molluscan shell evolution with review of shell calcification hypothesis. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **154**, 351–371 (2009).
67. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: An analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).
68. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25. <https://doi.org/10.1186/gb-2009-10-3-r25> (2009).
69. Stam, P. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. *Plant J.* **3**, 739–744 (1993).
70. Wu, Y., Close, T. J. & Lonardi, S. Accurate construction of consensus genetic maps via integer linear programming. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **8**, 381–394 (2011).
71. de Keyser, E., Shu, Q. Y., van Bockstaele, E. & de Riek, J. Multipoint-likelihood maximization mapping on 4 segregating populations to achieve an integrated framework map for QTL analysis in pot azalea (*Rhododendron simsii* hybrids). *BMC Mol. Biol.* **11**, 1. <https://doi.org/10.1186/1471-2199-11-1> (2010).
72. Martínez-García, P. J. *et al.* Combination of multipoint maximum likelihood (MML) and regression mapping algorithms to construct a high-density genetic linkage map for loblolly pine (*Pinus taeda* L.). *Tree Genet. Genomes* **9**, 1529–1535 (2013).
73. Peng, W. *et al.* An ultra-high density linkage map and QTL mapping for sex and growth-related traits of common carp (*Cyprinus carpio*). *Sci. Rep.* **6**, 26693. <https://doi.org/10.1038/srep26693> (2016).
74. Van Ooijen, J. W. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet. Res.* **93**, 343–349 (2011).
75. Voorrips, R. E. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **93**, 77–78 (2002).
76. Ouellette, L. A., Reid, R. W., Blanchard, S. G. & Brouwer, C. R. LinkageMapView - Rendering high resolution linkage and QTL maps. *Bioinformatics* **34**, 306–307 (2017).
77. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7. <https://doi.org/10.1186/s13742-015-0047-8> (2015).
78. Yu, J. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**, 203–208 (2006).
79. Yin, L. *et al.* rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genom. Proteom. Bioinform.* **19**, 619–628 (2021).
80. R Core Team R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. <https://www.R-project.org/> (2022).
81. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).
82. Kang, H. M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–1723 (2008).
83. Turner, S. D. qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Sour. Softw.* **3**, 731. <https://doi.org/10.21105/joss.00731> (2018).

Acknowledgements

The research leading to these results has received funding from the Interreg Atlantic Area Programme through the European Regional Development Fund for the project Co-Operation for Restoring Cockle Shellfisheries and its Ecosystem Services in the Atlantic Area (COCKLES, EAPA_458/2016; www.cockles-project.eu). Authors wish to thank L. Insua, S. Sánchez-Darriba for their technical contribution, and to all participants in the COCKLE's project for their support and useful comments. Alicia L Bruzos was supported by a predoctoral fellowship from the Spanish Ministry of Economy, Industry and Competitiveness (BES2016/078166). SCUBA CANCERS is funded by European Research Council (ERC) Starting Grant 716290 of Jose Tubio. Bioinformatic analysis was supported by Centro de Supercomputación de Galicia (CESGA). DR is supported by BBSRC Institute Strategic Funding Grants to the Roslin Institute (BBS/E/D/20002172, BBS/E/D/30002275, and BBS/E/D/10002070).

Author contributions

M.H. conducted all analysis and drafted the manuscript. D.R. collaborated in the analysis and the experiment design. B.G.P. constructed 2b-RAD libraries for map construction and GWAS. S.D. was involved in the experimental crosses, sampling, laboratory analysis and phenotyping. D.C. performed the experimental crosses at hatchery. A.L.B. was involved in sampling and gene mining. A.B. designed custom scripts and supervised the bioinformatic analysis. P.M. conceived the study, supervised the project and revised the manuscript. All authors collaborated in the manuscript and approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-21214-3>.

Correspondence and requests for materials should be addressed to P.M.

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

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



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

© The Author(s) 2022