Identifying genes associated with genetic control of color polymorphism in the pearl oyster Pinctada margaritifera var. cumingii (Linnaeus 1758) using a comparative whole genome pool-sequencing approach

Pierre-louis Stenger, Chin-Long Ky, Jeremie Vidal-Dupiol, Serge Planes, Céline Reisser

Identifying genes associated with genetic control of color polymorphism in the pearl oyster *Pinctada margaritifera* var. *cumingii* (Linnaeus 1758) using a comparative whole genome pool-sequencing approach

Pierre-Louis Stenger[1], Chin-Long Ky[1,2], Jeremie Vidal-Dupiol[2], Serge Planes[3] and Céline Reisser*[1,4]

1. Ifremer, IRD, Institut Louis-Malardé, Univ Polynésie française, EIO, F-98719 Taravao, Tahiti, Polynésie Française, France.
2. IHPE, Univ. Montpellier, CNRS, Ifremer, Univ. Perpignan Via Domitia, Montpellier France
3. PSL Research University, EPHE-UPVD-CNRS, USR 3278 CRIOBE, Labex Corail, Université de Perpignan, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France.
4. MARBEC, Univ Montpellier, CNRS, IFREMER, IRD, Montpellier, France

* Corresponding author: celine.reisser@ifremer.fr

Abstract

For hundreds of years, the color diversity of Mollusca shells has been a topic of interest for humanity. However, the genetic control underlying color expression is still poorly understood in mollusks. The pearl oyster *Pinctada margaritifera* is increasingly becoming a biological model to study this process due to its ability to produce a large range of colors. Previous breeding experiments demonstrated that color phenotypes were partly under genetic control, and while a few genes were found in comparative transcriptomics and epigenetic experiments, genetic variants associated to the phenotypes have not yet been investigated. Here, we used a pooled-sequencing approach on 172 individuals to investigate color-associated variants on three color phenotypes of economic interest for pearl farming, in three wild and one hatchery populations. While our results uncovered SNPs targeting pigment-related genes already identified in previous studies, such as PBGD, tyrosinases, GST, or FECH, we also identified new colorrelated genes occurring in the same pathways, like CYP4F8, CYP3A4 and CYP2R1. Moreover, we identified new genes involved in novel pathways unknown to be involved in shell coloration for *P. margaritifera*, like the carotenoid pathway, BCO1. These findings are essential to possibly implement future breeding programs focused on individual selection for specific color production in pearl oysters and improve the footprint of perliculture on Polynesian lagoon by producing less, but with a better quality.

Keywords

Pinctada margaritifera, Pooled Whole Genome Sequencing, color, pearl farming, pigmentation evolution, population genomics.

Introduction

The color diversity of Mollusca shells has been a topic of interest for scientists since hundreds of years (Williams, 2016). Mollusca is the largest phylum in the marine ecosystem which explains that, ultimately, only a small portion of its species has been studied. Consequently, the evolution of their widely diverse colors and shapes is poorly understood. Species with economical interest have historically been used by scientists to understand fundamental biological mechanisms such as color variation in bivalves.

Marine bivalves, like the scallops *Patinopecten yessoensis* (Jay, 1857) (Chang et al., 2007; Ding et al., 2015; Sun et al., 2015; Yuan et al., 2021), *Argopecten irradians* (Lamarck, 1819) (Adamkewicz & Castgna, 1988; Du et al., 2017) or the Pacific oyster *Crassostrea gigas*

(Thunberg, 1793) (Aguilera et al., 2014; Feng et al., 2018; Feng et al., 2015; Song et al., 2017; Bonnard et al., 2020; Hu et al., 2021, Bonnard et al. 2021) have already received much attention for the elucidation of their pigmentation pathways. More recently, the black-lipped pearl oyster *Pinctada margaritifera* var. *cumingii* (Linnaeus, 1758) emerged as a model to study pigmentation pathways (Ky et al., 2016; Ky et al., 2017a; Ky et al., 2018; Stenger et al., 2019 ; Stenger et al., 2021a; Stenger et al., 2021b). While color has been studied in scallops and oysters because of the importance of their appearance and visual flesh and shell quality for the food industry (Ding et al., 2015), color of the blacklipped pearl oyster inner shell is of importance for pearl production. *P. margaritifera* has the remarkable ability to produce one of the largest ranges of colored pearls for a marine species (Ky et al., 2014). This makes the Tahitian pearls unique and places pearl farming as the second economic resource in French Polynesia (Bouzerand, 2018). However, the lack of knowledge about how pearl quality is determined in the oyster, how it can be controlled, as well as an increasing occurrence of bad quality pearls on the market has led to the collapse of the pearl's economic value since 2001 (Bouzerand, 2018). To remedy to the crisis, a new policy aiming at producing less while increasing the quality has been enforced by local authorities, and research programs on the study of pearl quality trait determination and control began (Stenger et al., 2019; Latchere et al., 2018).

Traditional approaches to identify the proportion of genetic control of a trait involves experimental crossings between phenotypes (Zheng et al., 2013) or selffertilized mattings (Adamkewicz & Castgna, 1988) (or a combination of both techniques) (Kobayashi et al., 2004) to produce F1 and F2 generations and observe the segregation of a trait of interest while making assumptions about the underlying genetic control. These techniques have been used for the bluish vs. greenish strains of the Pacific abalone *Haliotis discus hannai* (Kobayashi et al., 2004), and in the orange vs. yellow bay scallops *Argopecten irradians* strains (Adamkewicz & Castgna, 1988), for which the shell color is determined by a similar one-locus two-allele system in both species. In the Chilean scallop *Argopecten purpuratus,* the purple, brown, orange, yellow, and white color strains were determined by two loci with a simple dominant model of epistasis (Winkler et al., 2001). In the noble scallop *Chlamys nobilis* (Reeve, 1852), the orange–purple, purple, and brown colors are explained by a one-locus three-alleles model (Zheng et al., 2013). In the

blacklipped pearl oyster *P. margaritifera*, a "one-locus three-alleles'' model with dominance was proposed to control external shell color, whereby the black allele is dominant to the red allele, which itself is dominant to the white "albino" allele (Ky et al., 2016). For the inner color of the shell, experimental crosses between red males and red females resulted in close to 100% red F1 (Ky et al., 2016), whereas crosses between yellow males and yellow females or between green males and green females gave both around 50% of yellow and 50% of green F1 (Ky, pers. com.). These results suggest that the red phenotype is likely controlled by few genes, with high impact alleles, whereas the control of yellow and green phenotype likely depends on multiple genes with low impact alleles (polygenic). Recently, transcriptomic (Stenger et al., 2021a) and epigenomic (Stenger et al., 2021b) studies have for the first time identified pathways leading to color variability in this species, as well as some key genes. However, putative genomic variants that could control those phenotypes have not yet been described.

Genetic variants involved in color phenotypes have been extensively studied in vertebrates. In felines, the acquisition of a stop codon in the tyrosinase related protein 1 (TYRP1) gene through a substitution from T to C at position 298 lead to a cinnamon phenotype (light brown color coat) (Lyons et al., 2005). A missense mutation is responsible for the replacement of the 618 arginine with a cysteine in the premelanosome protein 17 (PMEL17) gene, changing its 3D conformation and leading to the silver coat color (the uncommon silver dapple phenotype) in horses (Brunberg et al., 2006). SNPs that are located directly upstream, downstream or within intronic regions can also control color phenotypes by modulating gene expression. According to Hart et al 2013, an upstream SNP of the Solute Carrier Family 2 Member 4 gene (SLC2A4) produces blue eyes and light skin in humans (Hart et al., 2013). A single SNP in the intron of the HECT and RLD Domain Containing E3 Ubiquitin Protein Ligase 2 gene (HERC2) at position 86 determines the human blue-brown eye color (Sturm et al., 2008). Several other studies have also demonstrated that a single SNP in an intron can modify the expression of neighboring genes (Shima et al., 2006; Tokuhiro et al., 2003). Moreover, SNPs contained in intergenic regions can also impact phenotypic traits through their proximity to long intergenic noncoding RNAs (LincRNA) (Hangauer et al., 2013). In the Pacific oyster *Crassotrea gigas*, LincRNAs affect the expression of the pigment related genes tyrosinaselike proteins, dopamine, β-monooxygenase, chorion peroxidase and cytochrome P450 2U1, leading to different color phenotypes (Feng et al., 2018).

Since the rise of next generation sequencing (NGS) in the last decade, it is now possible to sequence entire genomes and identify color-associated SNPs. Genome-wide association studies (GWAS) with shell coloration were carried out for *Cepaea nemoralis*

(Linnaeus, 1758) (Richards et al., 2013), *Patinopecten yessoensis* (Ding et al., 2015) and *Argopecten irradians* (Du et al., 2017). For instance, in *P. yessoensis,* 395,646 SNPs were associated with the 'Ivory' strain and 310,649 with the 'Maple' strain (Ding et al., 2015). Among them, several SNPs impacted genes involved in metal transport, like the metalloreductase STEAP2 (Ding et al., 2015). In *A. irradians*, association analysis indicated that 126 SNPs were associated with carotenoid accumulation in the adductor muscle, like the cytochrome P450 family (Du et al., 2017). Whole-genome sequencing of separate individuals remains costly for an association approach, especially for non-model species, and the number of sequenced individuals is often low. A way to use more individuals at a reduced cost is to perform reduced-representation sequencing (like RAD-seq). However, the small proportion of the genome that is sequenced makes it easy for polygenic associated variants to not be detected. Sequencing pools of individuals (Pool-seq) represent another way to include many individuals while keeping the cost of sequencing down. It has been successful in identifying 17 SNPs linked to different color phenotypes in natural *Drosophila* populations (Bastide et al., 2013) and thus represents a costeffective technique to assess the association of variants to phenotypes by comparing allelic frequencies among those phenotypes. Moreover, this technique provides a great compromise between the sequencing costs (lower), the number of individuals sequenced (higher)and the coverage of the genome (whole genome).

The main objective of our study was to identify candidate SNPs associated with three inner shell color phenotypes using a total of 172 *P. margaritifera* individuals. For this, we used a pool-seq strategy on three *P. margaritifera* inner shell colors of interest for pearl farming: red, yellow and green phenotypes. The main questions were: (i) are there color associated SNPs to be identified in the red, green and yellow phenotype, (ii) could these SNPs impact known pigmentation genes or pathways, and (iii) can we identify novel genes or pathways associated to color that were not identified in previous studies?

Materials and methods

Samples collection

In this study, we used three wild populations (Takapoto, Katiu and Gambier) and one hatchery population selected for their strongly colored phenotypes and one hatchery

produced population, and selected individuals from the three phenotypes in each population (see map in Figure 1). All individuals had a black mantle phenotype. The map was obtained with the R package *Marmap* V. 1.0.3 (Pante & Simon-Bouhet, 2013; Pante & Simon-Bouhet, 2019) with the NOAA's (National Oceanic and Atmospheric Administration of U.S.A.) bathymetry data. The wild individuals were collected from Takapoto (14°37'00.5"S 145°12'11.4"W) and Katiu (16°25'59.1"S 144°21'11.9"W) atolls in the Tuamotu archipelago and Mangareva (23°06'55.8"S 134°59'03.6"W) island in the Gambier archipelago.

For each location and each color phenotype, 50 individuals were selected by visual appreciation of their color and phenotypically confirmed later by analyzing their inner shell color with *ImaginR* V 2.0 R package (Stenger, 2017; Stenger et al., 2019) (see "Color typing and analysis" section), resulting in a total of 450 individuals for the wild group. An additional 20 samples per color phenotype were selected from a hatchery color breeding program (total N=60). These F1 individuals originate from three multi-parental families selected for their highly saturated color phenotypes (10 males and 10 females from each identified phenotype sampled from the Gambier archipelago contributed to production of the progeny in the multi-parental families through mass spawning). These F1 individuals were used to enhance the genetic signal of putative allelic variants linked with color phenotype, by reducing the genetic noise due to genetic diversity in the dataset. On the other hand, wild populations are used to avoid any family effect and false positive association due to a possibly reduced genetic diversity in F1. As such, a total of 510 samples were used for DNA extraction.

All individuals were sampled in 2017 and had a dorso-ventral size between 8 and 12 centimeters for an estimated age of 2-3 years (age of maximum expression of their color phenotype; Ky, personal communication). For each individual, a 5 mm^3 piece of mantle was sampled and stored in ethanol 96% at 4°C. After 48 hours, ethanol was renewed. The shells of all individuals were cleaned and kept for inner shell color analysis.

Color typing and analysis

Shells of the sampled individuals were photographed with a Canon® PowerShot G9 placed in a Packshot Creator™ (V. 3.0.3.8), in order to prevent dark shadows and light reflection on the shell surface, as described in Stenger et al., (2019). Image processing and color qualification were realized in *ImaginR*, as described in Stenger et al., (2019). Using the HSV (Hue Saturation Value) color space, "ImaginR" characterizes inner shell color

variations using a machine learning protocol with a training dataset. The training dataset allows for building a distribution of HSV values that corresponds to the different color phenotypes (giving the software pictures of what is considered red, yellow or green phenotype in our species). The three HSV distributions of our species did not overlap, allowing for perfect identification. To phenotype a new individual, the picture of its shell is processed by the program, and its HSV values extracted and compared to the distribution of the training dataset. to reassign shells (and pearls) to pre-defined phenotypic categories constructed from the training set. Any HSV combination falling outside of the profiles of the 3 phenotypes will be classified as "unknown", while the rest of the individuals will be assigned to one of the three phenotypes. Individuals presenting both very high DNA quality and a validated affiliation to the red, yellow or green phenotype categories in *ImaginR* were kept for sequencing. Pairwise t-tests were conducted between phenotypes and sites for the HSV values, to test for any differences among geographic sites.

DNA extraction, QC, and pooling of samples

Samples were cleaned with DEPC before DNA extraction. Total DNA of 25mg of mantle was extracted for each individual with a QIAamp DNA Mini Kit from QIAGEN® (Cat No./ID: 51306) following manufacturer recommendations, with an RNase A treatment step (see manual). An additional step was added after overnight digestion in order to remove mucopolysaccharides following the protocol in Sokolov (2000): 50µL of a saturated KCl solution (34g KCl / 100g H_2O) was added to the digested samples, and the mixed solution was then centrifuged at 14,000 G for 15 minutes. DNA quality and quantity was assessed with a NanoDrop™ 1000 and agarose gels (500mL of agarose 1.2%, TBE 1X, SYBR® Safe 3µL (Invitrogen - Cat No./ID: S33102), 50 volts). Only high-quality DNA extraction were added to the pools. Pools were constructed using equimolar proportions of each individual of each phenotype and each sampling location. Figure 1 presents the number of sequenced individuals by pools according to their location and their phenotypes.

Sequencing approach

In total, 172 individuals were pooled according to their color and their population of origin, resulting in 12 pools (Figure 1). DNA libraries were prepared by Genome Québec

(MPS Canada) facilities, using a TruSeq DNA PCR Free for Illumina Kit for DNA libraries / Shotgun PCR Free (800ng DNA for library construction, Illumina TruSeq LT). Samples were sequenced on 12 lanes of an Illumina HiSeqX by the Genome Québec facilities (MPS Canada (one pool per lane).

Read trimming, mapping and SNP calling

Analyses were performed on the Ifremer Datarmor cluster. Raw reads quality was assessed with FastQC (Andrews, 2010). Reads were cleaned and adaptors were removed with Trimmomatic (Bolger et al., 2014) (V. 0.36 – illuminaclip 2:30:10; leading 28; trailing; 28 and minlen 40). BWA-MEM V. 0.7.15 (Li & Durbin, 2009) was used to map reads against *P. margaritifera*'s genome assembly (Reisser et al., 2020). The 12 BAM files were filtered with SAMtools V. 1.4.1 (Li et al., 2009) to keep only uniquely mapped and properly paired reads with a mapping quality score higher than 5, and then sorted and indexed. PCR duplicates were removed using the Picard MarkDuplicates (V. 1.119), and reads were corrected with SplitNCigarReads from GATK (McCormick et al., 2015). BAM files were then indexed with SAMtools and checked with the flagstat tools.

Variant calling was performed with Freebayes V. 1.2.0 (Garrison & Marth, 2012). The pooled-discrete and cnv-map parameters allowed us to precise the number of individuals (N individuals times ploidy) used for each pool in order to adapt the SNP calling to the sample size. VCF Filter V.1.0.0 (Müller et al., 2017) was used to filter out SNPs with a depth below 20. We used VCFtools V. 0.1.14 (Danecek et al., 2011) to filter out SNP loci with a minor allele frequency below 1% (remove rare alleles) and with any missing genotyping. BCFtools V. 1.4.1 (Narasimhan et al., 2016) was used to keep only biallelic SNPs loci. 22,169,780 SNPs were obtained at this step. Complex events reported by Freebayes (biallelic block substitutions) were decomplexed with *vt* and the decompose_blocksub algorithm (Tan et al., 2015), thereby recovering 1848 additional SNPs.

Multivariate analysis

Pi diversity was calculated using VCFtools V. 0.1.14 (Danecek et al., 2011) with a window of 10000 sites for each pool. A Kruskal-Wallis test and a pairwise Wilcoxon rank sum test were proceeded. For each pool, allelic frequencies (AF) from the VCF file were obtained by extracting the ratio of AD/DP for each of the SNPs.

To visualize relationships among samples, a principal component analysis was performed with the R packages *stats* V.3.6.1 (R Core Team) and *ggplot2* V.3.2.0 (Wickham, 2016) using AF data for each of the 12 pools.

Statistical models for association study

We used a combination of general linear model approaches to test for the presence of any significant differences in AF among color phenotypes, while controlling for geographical site of origin. Methodology for the combination of these statistical models were found in Hsu et al., 1996; Searle & Gruber, 1971; Hothorn et al., 2008 and Crawley, 2012.

For the GLM, the AF values were used as the response variable, while site and color values were used as explanatory variables.

for (i) SNP:

mod glm(AF ~ Color + Site)

Since there are two explanatory variables (color and site), the information was decomposed with another model because we were interested in comparing the levels of the two factors simultaneously. As such, a general linear hypothesis (GLHT) and multiple comparisons for parametric models (MCP) with ANOVA and post hoc Tuckey was implemented in a two-step strategy, separating color and site testing.

The GLHT-MCP was first used on the color results. The input model of the GLHTMCP was the result of the first GLM, and an ANOVA-Tuckey test was performed for the MCP results. Results were stored in an R variable and the same process was used for the site variable and stored into another R object.

To identify SNPs related to both color and geography to remove them from the results (because we are here interested in the universal control of color), we performed an additional test that encompasses all the results: simultaneous comparison of the levels of each factor was done using a new simple GLHT on the first GLM with specification of the linear hypotheses of both GLHT-MCP performed on color (K1) and on site (K2):

glht(mod, linfct = rbind(K1, K2))

A Bonferroni correction method (Hommel 1988) was applied on all *P* values. All analyses previously cited were performed with the R packages *multcomp* V. 1.4-10 (Hothorn et al., 2019), *readr* V. 1.3.1 (Wickham et al., 2022a), *dplyr* V. 0.8.3 (Wickham et al., 2022b), *tidyr* (Wickham & Girlich, 2022) and *stringr* V. 1.4.0 (Wickham, 2019). Parallelization of the

calculations was achieved with the R packages *foreach* V. 1.4.4 and *doParallel* V. 1.0.14 following Weston & Calaway, 2019 recommendations. This model combination allowed us to obtain *P* values for all site comparisons and color phenotype comparisons. A SNP was deemed significantly associated with color if it had a Bonferroni corrected *P* value < $1e10^{-9}$ in at least one pairwise color comparison and no geographic pairwise comparisons. In addition, a significantly associated SNP was deemed "color specific" if its association was found in all the pairwise comparison involving a specific color (eg. a green specific SNP will have a significant *P* value for both the green vs yellow comparison, and for the green vs red comparison).

The R package *circlize* (Gu, 2019) (V. 0.4.6) was used to create a chord diagram to visualize the color-associated SNPs. Histograms for allelic frequency distributions were obtained with ggplot2 with 30 bins. Pirate plots (raw data, descriptive statistics, and inferential statistics plotted in the same graphic from the *yarrr* R package V.0.1.5 (Phillips, 2017) were chosen to illustrate the differences between the twelve distributions. After checking data normality with a Shapiro test, a non-parametric Pairwise Wilcoxon test with a Bonferroni adjustment was performed with the R package *stats* to test any statistical differences between the distributions.

Functional analysis of associated SNPs

SnpEff and SnpSift V. 4.3 (Cingolani et al., 2012) were used to identify the potential impact of color-associated SNPs on *P. margaritifera*'s genes sequence. Visualization of these results were obtained with the R packages *ggplot2* and *cowplot* V1 (Claus & Wilke, 2019). The annotated VCF files containing the color associated SNPs are available as Supplementary File S01A, B and C.

We then continued the analysis keeping only color-associated SNPs with high, moderate and modifier impacts: high category includes impact like stop/start codon loss/gain, moderate category includes impacts like change in amino-acid, and modifier category includes impacts like intergenic region, intron, or upstream/downstream variant (2000bp around the gene, corresponding to the default parameter in SNPeff). GOATOOLS (Klopfenstein et al., 2018) was used to test for enrichment of GO terms in those SNPs, with Fisher's exact test (significance at p-value <0.05). Significantly enriched GO categories were visualized in REVIGO (Supek et al., 2011) for the biological process, cellular components and the molecular functions. The word clouds of the enriched GO

terms were obtained with the OmicsBox software (biobam) using the top 40 most enriched GO terms (see Supplementary File S02).

Amino acid changes and protein modeling

For all associated SNPs occurring in exonic regions of genes, complete exonic sequences were extracted using the *intersect* tool from BEDTOOLS (Quinlan & Hall, 2010). The corresponding exon sequence with the SNP location was translated into amino acids sequences with the Expasy web interface (https://www.expasy.org). The obtained amino acid sequences of these alternative proteins (bearing the associated variants) were compared with *P. margaritifera*'s protein sequences and the corresponding frame in Expasy were selected. Protein modeling for the reference and alternative sequences were obtained for *P. margaritifera's* PBGD (porphobilinogen deaminase), shematrin, cytochrome P450 3A29 and Cadherin-23 isoform X1 (see Stenger et al., 2021a) : i) sequences alignment were obtained with ClustalW for the *P. margaritifera* genome reference sequence and the alternative sequence carrying the amino-acid modification and (if any) with other species referential sequences, ii) I-Tasser (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) was used for secondary structure prediction for both reference and alternative *P. margaritifera*'s amino-acids sequences and for finding crystal structure (PDB) or the ligand binding site, iii) super-imposition of the Protein Database files was performed on the website tool SuperPose (http://superpose.wishartlab.com) with default parameters.

Results

Color typing

Inner shell color analysis of the wild individuals revealed that 87.36 %, 86.96 %, and 98.3 % of the sampled individuals were successfully assigned to the red, yellow and green phenotype respectively. The remaining individuals had color patterns that did not allow proper phenotypic characterization (multiple colors on the shell, as described in Ky et al.

2017b) and were thus removed from further analysis (see Figure 1 for details on sampling design).

Within a given color phenotype, the average hue value obtained with *ImaginR* did not significantly differ among sites (red phenotype 0.032, 0.032, 0.031, 0.036; green phenotype: 0.490, 0.500, 0.490, 0.500; yellow phenotype: 0.220, 0.210, 0.210, 0.180 for Takapoto, Katiu, Gambier and hatchery respectively). The darkness parameter of the yellow pool for the hatchery population is significantly different from Gambier (T test P= 0.011) but not from Tuamotu.

Bioinformatics analysis

Between 875,890,480 and 978,499,888 raw sequences were obtained for the twelve pools (see Table 1 for lane specific results). After trimming, between 8.46 % and 16.08 % sequences were removed and between 73.08 % and 75.66 % sequences were properly mapped and paired on the reference genome. After filtering steps (see Material and Methods for details) 22,171,628 SNPs were retained for statistical analysis.

Multivariate analysis, and genetic diversity

The Principal Component Analysis performed on allelic frequencies (AF) for the 12 pools preferentially groups the samples by geographical sites and not by color phenotypes (Figure 2): Takapoto and Katiu (the two atolls from the Tuamotu archipelago) are grouped together, while the Gambier and hatchery pools (the hatchery population originating from Gambier crossings) form another group. One exception to that is the yellow phenotype from the hatchery pool, which is differentiated from both the Tuamotu and the Gambier samples and represents an outlier.

Pi diversity does not statistically differ among the pools (see Supplementary Files S03A-B), or among the samples grouped according to geographical sites or color phenotypes.

Association study

We identified between 15326 and 32440 significantly associated SNPs depending on the color phenotype and geographical site considered (Figure 3). As expected, there

were more SNPs found in association with the geographical origin (from 21672 to 32440 SNPs) than with color phenotype (from 15326 to 16479 SNPs). Of these, 3622 SNPs were specifically associated with the red phenotype, 3564 to the yellow, and 3240 to the green (10426 SNPs in total). These color-associated SNPs are delimited by a black border in the chords of the diagram on Figure 3.

AF distributions of the color associated SNPs are shown in Figure 4. All distributions are skewed toward the alternative allele (higher homozygosity for the alternate allele), while the distribution for the 22,171,628 SNPs (associated and nonassociated) is skewed toward the reference allele (higher homozygosity for the reference allele). Pairwise Wilcoxon test showed that the AF distribution among red pools is more homogeneous than that among the yellow or among the green pools themselves. Moreover, the statistical differences of the Pairwise Wilcoxon test between AF distributions (Figure 4) didn't correlate with the differences in hue, saturation or darkness values between sites and phenotypes previously identified through the pairwise t-test.

Functional analysis of color-associated SNPs

Figure 5 shows the number of SNPs classified by their impact for each comparison. The modifier category contained the highest number of SNPs (97.61 % of the total number of effects on average). The moderate and high categories only represented 1.05 % and 0.09% of the total number of effects respectively. Table 2 shows the most enriched GO terms according to GOATOOLS for each three phenotypes. The three color phenotypes share 12 enriched GO categories among which 3 were common to the red and green, 3 between red and yellow and 4 between yellow and green. Interestingly, there were 31 color-specific enriched GO terms: 11 GO terms were specifically enriched in the red phenotype, as for example "hydroxymethylbilane synthase activity" (GO:0004418), 11 GO terms were specifically enriched in the yellow phenotype, such as "calcium ion binding" (GO:0005509), and 9 GO terms were specifically enriched in the green phenotype like "heterocyclic compound binding" (GO:1901363) (Table 2).

Pigmentation pathways associated SNPs

Of the 10426 color-specific associated SNPs (see Supplementary File S01A, B and C), only 1.5% had GO categories related to pigment trafficking and biomineralization processes, like encapsulation of molecules, vacuole trafficking pigments, or biomineralization. Here we focused on SNPs impacting genes directly involved in pigmentation (these SNPs can be found in Supplementary Files S04 A, B, and C). A total of 54 different pigmentation related proteins encoded by 108 different transcripts were impacted by SNPs (Table 3 A, B and C). Red specific SNPs impacted 18 different genes encoding various pigment-related proteins, like β, β -Carotene 15,15'-dioxygenase gene, porphobilinogen deaminase and cytochrome P450 2C8 (Table 3). Yellow-specific SNPs impacted, 66 different pigments related genes, such as Xanthine dehydrogenase/oxidase, Cytochrome P450 3A11-like or many Copper/zinc superoxide dismutases (Table 3). Green-specific SNPs impacted 27 pigment-related genes, such as flavin reductase (NADPH), glutathione S-transferase 1-like, or three cytochrome P450 2C8-like proteins (Table 3). Other genes were impacted by different SNPs in two color phenotypes: cytochrome P450 4F8-like and the visual pigment receptor peropsin-like (RRH) genes were impacted by different SNPs identified in yellow and red phenotypes. Cytochrome P450 2C8-like was also impacted by different SNPs in green and yellow phenotypes (Table 3).

Amino acid changes and protein modeling

Among the 108 pigment-related genes impacted by SNPs, only four had SNP impacting exonic sequences: a porphobilinogen deaminase (PBGD) sequence (scaffold2460|size144317.10), a shematrin sequence (scaffold2460|size144317.17), a cytochrome P450 3A29 (CYP3A29) (scaffold2294|size169326.2) and a Cadherin-23 isoform X1 (scaffold5945|size55440.1). The SNP impacting the PBGD sequence induced a change from a 307 threonine into a 307 isoleucine; the SNP impacting the shematrin sequence induced a change from a 340 aspartic acid to a 340 asparagine; the SNP impacting the CYP3A29 induced a change from a 140 leucine to a 140 phenylalanine and the SNP impacting the Cadherin induced a change from a 119 threonine into a 119 alanine.

Alignment of the human reference PBGD (Song et al., 2009) sequence with the *P.*

margaritifera genome's reference sequence and the alternative one revealed that both *P. margaritifera* sequences show a 31.58 % similarity to the human PBGD (Figure 6A). The presence of the SNP reduces the similarity between the two *P. margaritifera* sequences at 98.96 %. The secondary structure prediction by I-Tasser showed that the AA substitution leads to the folding of a beta strand at the C terminal part of the alternative *P. margaritifera*'s PBGD sequence (see Supplementary File S05A and B). The superimposition between the reference and the alternative pearl oyster PBGD has a score of 488.8 (needle program; Figure 6A), indicating significant 3D conformation differences.

Since there is no shematrin reference protein in the Protein Data Bank, no match with shematrin protein was found with our *de novo* pearl oyster shematrin PDB. The comparison of the secondary structure of both reference and alternative *P. margaritifera*'s shematrin sequences shows that the change from 340 aspartic acid to 340 asparagine in the alternative sequence induces the folding of an alpha helix, from position 349 to 350 (Figure 6B). The protein modeling visualization indicates that the reference and alternative sequences are very close to each other at the tertiary structure level (Figure 6B, Supplementary File S05C and D). The superimposition needle score for both pearl oyster shematrins is relatively high, with a value of 1959.0.

For CYP3A29, the protein modeling of reference and alternative pearl oyster CYP3A29 highlight structural differences between them (Supplementary File S05E and F). While the reference sequence displays an alpha helix from the amino acid 138 to 151, the alternative sequence displays a beta strand from the amino acid 138 to 142 followed by an alpha helix from the amino acid 143 to 151 (Figure 6C). Despite the high needle score (1341.0) for the reference and alternative pearl oyster CYT3A29 sequences, the superimposition of their PDB is offset, due to the secondary structure changes linked to the SNP (Figure 6C).

Alignment of the *P. margaritifera*'s reference Cadherin-23 isoform X1 sequence and the alternative indicated that both sequences are very similar (99.72 %) while the alignment of the secondary structure shows a lower similarity (97.72 %) (Figure 6D, Supplementary File S05G and H). The superimposition of both reference and alternative pearl oyster cadherin-23 matches with a small needle score of 111.0 (Figure 6D), indicating a divergence in the conformation between the proteins.

Discussion

The major quality trait of the French Polynesian pearls relies on the very diverse color range of the inner shell of *P. margaritifera* var. *cumingii,* and as such of its cultured pearls, a diversity that cannot be found in other *Pinctada* species (Ky et al., 2013; Ky et al., 2017a). However, little is known about the putative genetic control of these colors in pearl oyster species (Lemer et al., 2015; Shinohara et al., 2018; Williams, 2016). Our aim was thus to identify SNPs associated to three different inner shell colors in *P. margaritifera* (red (Ky et al., 2015), yellow (Ky et al., 2015) and green (Ky et al., 2017a)) to better understand the pathways involved in pigment production and the associated genetic mutations that impact them. While we identified SNPs impacting a set of genes that had been previously identified as differentially expressed among the different color phenotypes (thus highlighting a possible genetic control of their expression), we also uncover previously unknown pathways impacted by color specific SNPs, implicating novel candidate genes for the synthesis of pigment and expression of color phenotype in the Tahitian pearl oyster.

When considering both the geographical origin and the color of oysters, our results showed that geographical origin was the main driver of genetic differentiation among our samples. Indeed, samples generally clustered by archipelago, except for the yellow individuals from hatchery which were outliers to all other samples. In wild populations, the yellow phenotype can present peculiarities regarding the positioning of the colored area of the inner shell: the colored area for the yellow individuals was indeed generally found at the bottom of the posterior section of the inner shell, while for the two other phenotypes, the colored spot was present as a large band in all posterior, connection, middle and anterior sections of the inner shell (Ky et al., 2018). Moreover, results from *ImaginR* color qualification showed that the yellow hatchery individuals present a specific hue compared to all other yellow phenotypes, while having a significantly different global darkness value than that of the Gambier individuals. While it suggests the existence of a peculiar pigmentation deposition for the yellow phenotype between hatchery's individuals and the others. It cannot be explained by a possible family effect or diversity effect, since allelic frequencies distribution (AFD) of the color associated SNPs indicated that the yellow hatchery AFD were similar to the other yellow AFD. In addition, levels of Pi diversity were found to be similar in all our samples, so that a possible inbreeding effect in the yellow hatchery sample is unlikely. Thus, the origin of the genetic variation and the outlier status of the yellow hatchery population cannot be identified at this stage but is to be noted.

Putative genetic control of the red phenotype

Our results demonstrated that the red phenotype might be controlled by a limited number of SNPs and might be less polygenic than the green and yellow phenotypes (Slatkin, 1978; Stahl et al., 2012). Indeed, the red phenotype had a very similar AF distribution in all the geographical sites, while the two other phenotypes showed more variation. This observation is also congruent with the phenotypic segregation observed in F1 experimental crosses between red individuals from previous studies (see introduction).

The red specific SNPs showed functional enrichment for hydroxymethylbilane synthase activity (GO:0004418), a pathway containing the gene porphobilinogen deaminase (PBGD). This GO category was already identified as specific to the red color in a previous study (Stenger et al., 2021a), in which PBGD was significantly downregulated in the red phenotype compared to the yellow and green phenotypes. This downregulation causes a form of acute intermittent porphyria (AIP) in mammals, linked with the accumulation of uroporphyrin I in cells, a well-known red pigment that was previously identified in marine bivalve shells (Stenger et al., 2021a). This kind of porphyria may here be the result of a non-functional PBGD protein due to a SNP (Balwani & Desnick, 2012; Meyer et al., 1972; Schneider-Yin et al., 2002; Siersema et al., 1990; Song et al., 2009). Five different genes encoding porphobilinogen deaminases were found to be impacted by 91 SNPs in the red phenotype, though only one was located within an exon. This SNP created a beta strand at the positions 92-93 of the protein and a 3D conformation change with the reference protein. However, we did not find any scientific literature on the cost of this change for the protein's functionality. The other SNPs were located in introns or upstream/downstream of the gene sequence, all of which are known to influence the expression levels of the neighboring genes (Shima et al., 2006; Tokuhiro et al., 2003). These SNPs could thus lead to a perturbation of the heme pathway, leading to an accumulation of the red pigment uroporphyrin I due to downregulation of the PBGD, in the red pearl oyster phenotype, as previously hypothesized (Stenger et al., 2021a), and validation by experimental crosses, phenotyping and genotyping will be necessary.

In addition to the previously identified heme pathway, our genome-wide study highlighted a new pigmentation pathway not previously identified as involved in the red phenotype: the carotenoid pathway. Carotenoids are well-known red, brown, and yellow pigments (Hornero-Méndez et al., 2000; Lado et al., 2015; Tolmach & Graham, 1942). A

SNP was located in the 5' region of the β, β -Carotene 15,15'-dioxygenase gene (BCO1) and has a higher frequency in the red phenotype than others. BC01 is a key enzyme in betacarotene metabolism to vitamin A, and SNPs located in 5' regions of genes involved in this pathway are known to impact the production of vitamin A (Lietz et al., 2012). Four other genes impacted by SNPs were found to intervene in carotenoids biosynthesis and in the correlated retinol metabolic pathway: three cytochromes P450 (CYP26A1, CYP2C8 and CYP3A29), impacted by 8 SNPs, and a UDP-glucuronosyltransferase 1-2 (UGT), a gene previously described as differentially regulated in different color phenotypes in pearl oysters (Stenger et al., 2021a). For the cytochrome P450 family, one SNP was located in an exonic region of CYP3A29. This SNP significantly changed the tertiary conformation of the protein by shortening an alpha helix and creating a new beta strand at the modified amino acid position, which suggests a possible impact on its functionality. SNPs affecting BCO1, CYP26A1, CYP2C8, CYP3A29 and UGT could affect concentrations of circulating carotenoids by modifying the degradation of β-carotene molecules into all-trans-retinoic acids, leading to an accumulation of β-carotene that would have to be excreted/degraded into pigmented carotenoids forms (Biesalski et al., 2007; Chichili et al., 2005; Hill & Johnson, 2012; Hornero-Méndez et al., 2000; Lado et al., 2015; Lietz et al., 2012; Nadin & Murray, 1999; Naveed et al., 2018; Tolmach & Graham, 1942). Recently, retinol and retinoic genes have been confirmed to be associated with shell color of *Crassostrea gigas* in the orange and black phenotype (Li et al., 2021).

The purine metabolism pathway also was impacted by a SNP upstream of the bifunctional purine biosynthesis protein (PurH - EC:2.1.2.3 3.5.4.10 in the red phenotype). PurH controls the synthesis of FAICAR molecule (5'-Phosphoribosyl-5-formamido-4imidazolecarboxamide) and can also use PurH as a substrate to produce IMP (Inosine 5'monophosphate), the final molecule in the de novo purine nucleotide pathway (Beardsley et al., 1998). IMP molecule is the direct precursor to Guanosine 5' triphosphate (GTP) (Ng et al., 2009) and was identified as differentially methylated and potentially involved in coloration process in the pearl oyster *P. margaritifera* (Stenger et al., 2021b). This specific GTP is involved in the synthesis of yellow pigments in the pterins pathways (xanthopterins and sepiapterins) and in pheomelanin or in eumelanin in the RaperManson pathway (Stenger et al., 2021a; Stenger et al., 2021b). Since IMP can be created by different ways, the pigment production in pterins and Raper-Mason pathway could be slightly reduced in the red phenotype. We also identified 13 SNPs related to the Shematrin-like protein 1. The superimposition of the reference and alternative shematrin show differential conformation in the C-terminal end of the protein suggesting potentially weaker functionality. Shematrins are Mollusca proteins expressed specifically in the edge

region of the mantle (Marie et al., 2012; Yano et al., 2006). These proteins could take part in the calcification of the prismatic layer of the pearl oyster shell (Yano et al., 2006), and were over-expressed in black strains compared to the albino strains in *P. margaritifera* (Lemer et al., 2015).

In conclusion, individuals of the red phenotype have specific SNPs impacting the PBGD genes, which could affect gene expression and lead to the accumulation of the red pigment uroporphyrin I in the shell, as described in previous work (Stenger et al., 2021a). They also have SNPs impacting pigments from the pterin pathway, a pathway previously identified in Stenger et al. (2021a, 2021b). However, the present results identified a new pathway that was not described before: indeed, we found multiple SNPs impacting genes of the carotenoid pathways.

Putative genetic control of the yellow phenotype

In the yellow phenotype, we identified 18 impacted genes involved in the melanin pathway, such as laccases and tyrosinases. The melanin pathway uses DOPA and DOPA quinone as precursors to synthesize eumelanin and pheomelanin. DOPA quinone can be used as a substrate by glutathione-S-transferase, which binds glutathione (Sonthalia et al., 2016) to produce glutathionyldopa, leading to pheomelanin (yellow to red color) instead of eumelanin (brown to black color) (Stenger et al., 2021a).

Several yellow-specific SNPs impacted the cytochrome P450 family. We identified two SNPs in two introns of cytochrome P450 3A (CYP3A). CYP3A4 metabolizes the anthocyanidin and anthocyanin pigments (Srovnalova et al., 2014), and was downregulated in yellow shells of *Pinctada fucata martensii* when compared to the darker shells (Xu et al., 2019). Three upstream and one intronic SNPs were also identified in CYP4F8, a gene that is downregulated in patients with lentigo (small reddish pigmented spot on the skin that induce freckles) compared to patients with normal skin (Shin, 2012). An upstream yellow specific SNP was found in a gene coding for vitamin D(3) 25hydroxylase (also known as cytochrome P450 2R1 – CYP2R1 - EC 1.14.14.24). CYP2R1 intervenes specifically in the steroid biosynthesis and controls the synthesis of calcidiol from vitamin D3. Calcidiol can become calcitriol in the next steps, which is the active form of the vitamin D. There is a huge literature about the role of vitamin D3 and skin coloration (Jablonski & Chaplin, 2010). Indeed, the downregulation of this gene is found in light skin types compared to darker skin in humans (Liu et al., 2015). The putative implication of this gene to the yellow *P. margaritifera* phenotype is further supported by the previous

transcriptomic and epigenetic analyses between yellow and green phenotype in *P. margaritifera* (Stenger et al., 2021a; Stenger et al., 2021b), which identified differential expression and methylation of many enzymes involved in steroids and in vitamins D3 pathways.

The purine salvage pathway (Sculley et al., 1992) was also impacted by yellow specific SNPs. Two SNPs were found in the hypoxanthine-guanine hosphoribosyltransferase gene (HGPRT) in the yellow phenotype. HGPRT plays a central role in the generation of purine nucleotides, and its deficiency leads to the Lesch-Nyhan syndrome characterized by the overproduction of sand-like crystals of uric acid (Sculley et al., 1992). Also, an upstream SNP impacting xanthine dehydrogenase/oxidase (XDH) was identified in the yellow phenotype. Xanthine dehydrogenase and related genes code for molybdenum-containing hydroxylases that catalyze purine and pteridine reactions (Brondino, Romão, Moura, & Moura, 2006). This family of genes contribute red, orange, and yellow pigmentation in many animals (Watt, 1972; Shamim et al., 2014; Pimsler et al., 2017). This gene encodes an enzyme that oxidizes the yellow pteridine xanthopterin compound to the colorless pteridine leucopterin compound in *Colias* butterflies (Watt, 1972).

Other classically described genes involved in biomineralization processes were impacted by yellow specific SNPs. Three upstream and one downstream SNPs were found in the vicinity of two genes encoding for an amorphous calcium carbonate binding protein 1 (ACCBP). This gene is involved in morphology of nacre lamellae in the shell of *Pinctada* species (Ma et al., 2007; Shi et al., 2013). Indeed, this gene could inhibit the growth of undesired aragonite crystal faces and ceasing the nucleation and growth of calcite (Ma et al., 2007). These could modify the morphology of nacre lamellae (Ma et al., 2007), and since such changes could modify the color by physical mechanisms (Rousseau & RollionBard, 2012; Stenger et al., 2021b), further studies on this gene could be interesting.

Finally, 38 different genes encoding putative copper/zinc superoxide dismutase (Cu/Zn SOD) were impacted by 47 yellow-specific SNPs. Pigmentation pathways usually generate highly reactive molecules (porphyrins for example) that can be oxidized and have the potential to create reactive oxygen species (ROS), deleterious to the cell. Cu/Zn SOD are enzymes that catalyze the dismutation of the superoxide radical into either ordinary molecular oxygen or hydrogen peroxide. For example, Cu/Zn SOD expression is incredibly high in patients with vitiligo because of the melanogenesis pathway perturbation (Wacewicz et al., 2018).

In conclusion, the yellow individuals have SNPs impacting laccases, tyrosinases, HGPRT and XDH, which could be the source of the differential expression identified in these genes previously (Stenger et al., 2021a). In addition, our results identified new genes (CYP4F8, CYP3A4 and CYP2R1) in these pathways.

Putative genetic control of the green phenotype

A previous transcriptomic study (Stenger et al., 2021a) found that the GST enzyme (see yellow phenotype above) which occurs in the melanin pathway is downregulated in the green phenotype compared to the yellow phenotype. GST uses dopaquinone to produce specifically pheomelanin, a yellow pigment. In the green phenotype, we found a green specific SNP in GST which could explain the downregulation of this enzyme detected in previous studies (Stenger et al., 2021a), leading to the preferential production of the black eumelanin pigment preferentially to the yellow pheomelanin pigment. Recently, the same hypothesis has been raised for the black *Pinctada fucata martensii* color line (Adzigbli et al., 2020). Moreover, it was proven recently that the regulation of expression of different tyrosinase genes (essential enzymes for melanism process) play a vital role in the shell color of the bivalve *Ruditapes philippinarum* (Jiang et al., 2020).

One green specific upstream SNP impacted the ferritin 1 gene. This enzyme (EC:1.16.3.2) stores iron in a soluble, non-toxic, readily available form and could be used in the heme biosynthesis pathway (Theil et al., 2012; Wang et al., 2011). Also, one green specific SNP impacted ferrochelatase (FECH). FECH intervenes in the latest steps of the heme biosynthesis, to produce heme from protoporphyrin IX. Also belonging to the heme pathway, flavin reductase NADPH (BLVRB gene - EC:1.3.1.24) was impacted by a green specific SNP. This enzyme converts biliverdin (green pigment) into bilirubin (yellow pigment) (Cunningham, Gore, & Mantle, 2000). The SNP impacting BLVRB could affect the synthesis of the protein, leading to green biliverdin accumulation in the green phenotype, as previously hypothesized (Stenger et al., 2021a).

Thirteen SNPs impacting nine different cadherin-23 genes (Figure 6B) were identified in our dataset, among which four SNPs impacted four green specific transcripts. Although cadherin-23 couldn't be directly linked to any green pigment, this protein is involved in several pigmentation processes in other organisms. Cadherin-23 overexpression has been associated with the "purple zigzag" phenotype in the clam *Meretrix*, with an up-regulation in this phenotype compared to the white phenotype (Yue et al., 2015).

Classic biomineralization genes were also identified as possibly impacted by SNPs. Four green specific SNPs were identified in up- and downstream regions of three perlucin-like genes. Perlucin is involved in nucleation and subsequent growth of calcium carbonate crystals (Blank et al., 2003). This protein allows the connection between the chitin and aragonite layers (Sun et al., 2015). Recently, this protein was recognized to be linked to the pigmentation of the shell of *Ruditapes philippinarum* (Nie et al., 2020). Also, multiple SNPs were identified in calcium-binding protein cml8 and calcium-binding protein LPS1-alpha like, which are potential calcium sensors found only in the green phenotype.

In conclusion, the green individuals have SNPs impacting the previously identified melanin pathway and the heme pathway. However, cadherins are new genes that were not previously described in color associated genes in *P. margaritifera*. These genes are impacted by several SNPs for the green phenotype and should be investigated in further studies.

Conclusion

The goal of the study was to look for genetic variants associated with three different color phenotypes of the Tahitian pearl oyster: green, yellow and red. First, we confirmed that the genomic control of color in *P. margaritifera* is polygenic, with SNPs impacting multiple genes involved in several pigmentation pathways, sometimes shared among the different colors. Moreover, we noted the fact that the red phenotype was likely less polygenic than yellow and green phenotype, since the allelic frequency variation of the associated SNPs among the 4 populations studied was not as strong as for the two other phenotypes. Of the 108 pigment-related SNPs we identified, four fell into exonic sequences, while many others were located into upstream or downstream regions, in intergenic regions, or in introns. The four exonic SNPs impacted the 3D conformation of the proteins they modified, but their actual effect on gene expression and the resulting phenotype will have to be confirmed by further experimental studies, as is the case of upstream and downstream SNPs. Associated SNPs located in intergenic regions cannot be overlooked either, as they could impact LincRNAs and control color phenotypes, like for *C. gigas*.

Our whole-genome pooled sequencing approach revealed that many genes impacted by color specific SNPs had been identified as candidate genes in the transcriptomic analysis of the same color phenotypes. However, our study highlighted complementary pathways not identified previously, and which are known to be involved

in color expression in other Mollusca and animals, such as the carotenoids pathway for the red phenotype (BCO1). The presence of these pigments in the shell of the pearl oysters will have to be confirmed through biochemistry, since the RAMAN spectroscopy analysis used in previous studies on the shells of the tahitian pearl oysters is not effective at identifying carotenoids. Also, new candidate genes were also identified for the yellow phenotype (CYP4F8, CYP3A4 and CYP2R1), and for the green phenotype (cadherin-23). Finally, we identified genes that are not directly involved in pigment synthesis but that are known to intervene in the control of potential cell damage due to the reactive nature of many pigments (oxidative stress and ROS degradation) like the Cu/Zn SOD genes.

The actual functional role of the candidate SNPs in pigmentation pathways must be validated. Currently, it is not possible to cultivate mantle cell lineages of *P. margaritifera* in the lab, to provide a way to use crispR techniques and observe the color of nacre precipitation of these cells. RNAi techniques were tested but are very tedious to use on such a species, and very stressful for the animal. Currently, our only perspective is to use these SNPs and genotype as many individuals as possible to validate the pattern of segregation we see here at a larger scale. A SNP Chip has been developed and includes all the SNPs identified in this study, to validate the association of these SNPs to color morphs. The chip is currently being tested in breeding programs with pearl farmers. Our results open new doors for the implementation of future breeding programs focused on individual genomic selection for specific color production in pearl oysters, and improve the footprint of perliculture on Polynesian lagoon by reducing the production volume while increasing the quality of the pearls.

ACKNOWLEDGEMENTS

This study was supported by grants from the "Direction des Ressources Marines", through the AmeliGEN project (# 10065/MEI/DRMM). The authors would like to thank the Regahiga Pearl farm (Mangareva Island, Gambier archipelago, French Polynesia) and the Pahai Poe pearl farm (Apataki atoll, Tuamotu archipelago, French Polynesia) for providing the pearl oysters used in this study. This study is set within the framework of the « Laboratoire d'Excellence (LabEx) » TULIP (ANR-10-LABX-41), and the LabEx CeMEB (ANR-10-LABX-04-01).

CONFLICT OF INTEREST

The authors have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

Raw reads for this manuscript are openly available at the Sextant Database https://sextant.ifremer.fr/record/e0a8f7bc-423d-4043-b997-d839a083f01b/ (Reisser, 2018).

The whole pipeline and code scripts can be found on GitHub (PLStenger/Pearl_Oyster_Colour_Population_Genomics/00_scripts).

Bibliography

Adamkewicz, L., & Castgna, M. (1988). Genetics of Shell Color and Pattern in the Bay Scallop *Argopecten irradians*. *The Journal of Heredity, 79(1)*, 14–17.

Adzigbli, L., Wang, Z., Li, J., & Deng, Y. (2020). Survival, retention rate and immunity of the black shell colored stocks of pearl oyster *Pinctada fucata martensii* after grafting operation. *Fish and Shellfish Immunology, 98*, 691–698. https://doi.org/10.1016/j.fsi.2019.11.003

Aguilera, F., McDougall, C., & Degnan, B. M. (2014). Evolution of the tyrosinase gene family in bivalve molluscs: Independent expansion of the mantle gene repertoire. *Acta* Biomaterialia, 10(9), 3855–3865. https://doi.org/10.1016/j.actbio.2014.03.031

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics. https://www.bioinformatics.babraham.ac.uk/projects/fastqc

Balwani, M., & Desnick, R. J. (2012). The porphyrias : advances in diagnosis and treatment. *ASH Education Program Book, 2012(1),* 19–27. https://doi.org/10.1182/asheducation2012.1.19

Bastide, H., Betancourt, A., Nolte, V., Tobler, R., Stöbe, P., Futschik, A., & Schlötterer, C. (2013). A genome-wide, fine-scale map of natural pigmentation variation in *Drosophila melanogaster. PLoS Genetics* 9 (6): e1003534. https://doi.org/10.1371/journal.pgen.1003534

Beardsley, G. P., Rayl, E. A., Gunn, K., Moroson, B. A., Seow, H., Anderson, K. S., & Davies, J. (1998). Structure and functional relationships in human purH. *In Purine and Pyrimidine Metabolism in Man IX* (Vol.43, pp. 221–226). https://doi.org/10.1007/978-1-4615-5381-6

Biesalski, H. K., Chichili, G. R., Frank, J., von Lintig, J., & Nohr, D. (2007). Conversion of β-Carotene to Retinal Pigment. *Vitamins and Hormones, 75(6),* 117–130. https://doi.org/10.1016/S0083-6729(06)75005-1

Blank, S., Arnoldi, M., Khoshnavaz, S., Treccani, L., Kuntz, M., Mann, K., Grathwohl, G., & Fritz, M. (2003). The nacre protein perlucin nucleates growth of calcium carbonate

crystals. *Journal of Microscopy, 212(3),* 280–291. https://doi.org/10.1111/j.1365-2818.2003.01263.x

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Genome analysis Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30(15),* 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bonnard, M., Cantel, S., Boury, B., Parot, I. (2020). Chemical evidence of rare porphyrins in purple shells of *Crassostrea gigas* oyster. *Scientific Reports, 10,* 12150. https://doi.org/10.1038/s41598-020-69133-5

Bonnard, M., Boury, B., Parot, I. (2021). First evidence of ommochromes in Bivalves: the case of the edible oyster *Crassostrea Gigas. Research Square Platform LLC*. https://doi.org/10.21203%2Frs.3.rs-677796%2Fv1

Bouzerand, E. (2018). Le bilan de la perle en 2016*. Institut de la statistique de la Polynésie Française, Papeete.* https://www.ispf.pf/content//uploads/Le_bilan_de_la_perle_en_2016_c99b1781ea.pdf

Brunberg, E., Andersson, L., Cothran, G., Sandberg, K., Mikko, S., & Lindgren, G. (2006). A missense mutation in PMEL17 is associated with the Silver coat color in the horse. *BMC Genetics, 7*, 1–10. https://doi.org/10.1186/1471-2156-7-46

Chang, Y. Q., Chen, X. X., Ding, J., Cao, X. B., Li, R. L., & Sun, X. W. (2007). Genetic diversity in five scallop populations of the Japanese scallop (*Patinopecten yessoensis*). *Acta Ecologica Sinica, 27(3),* 1145–1152.

Chichili, G. R., Nohr, D., Schäffer, M., Von Lintig, J., & Biesalski, H. K. (2005). β-carotene conversion into vitamin A in human retinal pigment epithelial cells*. Investigative Ophthalmology and Visual Science, 46(10),* 3562–3569. https://doi.org/10.1167/iovs.050089

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly, 6(2),* 80–92. https://doi.org/10.4161/fly.19695

Claus, A., & Wilke, O. (2019). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. https://CRAN.R-project.org/package=cowplot

Crawley, M. J. (2007). The R book. Chichester, England: Wiley.

Cunningham, O., Gore, M. G., & Mantle, T. J. (2000). Initial-rate kinetics of the flavin reductase reaction catalysed by human biliverdin-IXβ reductase (BVR-B). *Biochemical Journal, 345(2)*, 393–399. https://doi.org/10.1042/0264-6021:3450393

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics, 27(15),* 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Ding, J., Zhao, L., Chang, Y., Zhao, W., Du, Z., & Hao, Z. (2015). Transcriptome sequencing and characterization of Japanese scallop *Patinopecten yessoensis* from different shell color lines. *PLoS ONE, 10(2),* 1–18. https://doi.org/10.1371/journal.pone.0116406

Du, X., Song, K., Wang, J., Cong, R., Li, L., & Zhang, G. (2017). Draft genome and SNPs associated with carotenoid accumulation in adductor muscles of bay scallop (*Argopecten irradians*). *Journal of Genomics, 5,* 83–90. https://doi.org/10.7150/jgen.19146

Feng, D., Li, Q., Yu, H., Zhao, X., & Kong, L. (2015). Comparative transcriptome analysis of the pacific oyster *Crassostrea gigas* characterized by shell colors: Identification of genetic bases potentially involved in pigmentation. *PLoS ONE, 10(12),* 1–17. https://doi.org/10.1371/journal.pone.0145257

Feng, D., Li, Q., Yu, H., Kong, L., & Du, S. (2018). Transcriptional profiling of long non-coding RNAs in mantle of *Crassostrea gigas* and their association with shell pigmentation. *Scientific Reports, 8(1),* 1–10. https://doi.org/10.1038/s41598-018-19950-6

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing, 1–9. arXiv:1207.3907 [q-bio.GN]

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). “circlize implements and enhances circular visualization in R.” *Bioinformatics, 30,* 2811-2812.

Hangauer, M. J., Vaughn, I. W., & McManus, M. T. (2013). Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. *PLoS Genetics, 9(6)*. https://doi.org/10.1371/journal.pgen.1003569

Hart, K. L., Kimura, S. L., Mushailov, V., Budimlija, Z. M., Prinz, M., & Wurmbach, E. (2013). Improved eye- and skin-color prediction based on 8 SNPs. *Croatian Medical Journal, 54(3),* 248–256. https://doi.org/10.3325/cmj.2013.54.248

Hill, G. E., & Johnson, J. D. (2012). The vitamin A-Redox hypothesis: A biochemical basis for honest signaling via carotenoid pigmentation. *American Naturalist, 180(5).* https://doi.org/10.1086/667861

Hornero-Méndez, D., De Guevara, R. G. L., & Isabel Mínguez-Mosquera, M. (2000). Carotenoid biosynthesis changes in five red pepper (*Capsicum annuum L.*) cultivars during ripening. Cultivar selection for breeding. *Journal of Agricultural and Food Chemistry, 48(9)*, 3857–3864. https://doi.org/10.1021/jf991020r

Hothorn, Bretz, F., Westfall, P., Heiberger, R. M., Schuetzenmeister, A., & Scheibe, S. (2019). Multcomp package (Simultaneous Inference in General Parametric Models). Retrieved from http://multcomp.r-forge.r-project.org

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50(3)*, 346–363. https://doi.org/10.1002/bimj.200810425

Hsu, J. C., Nelson, B., Journal, S., Statistics, G., Mar, N., Hsu, J. C., & Nelson, B. (1996). Multiple Comparisons in the General Linear Model. *Journal of Computational and Graphical Studies, 7(1)*, 23–41.

Hu, B., Li, Q., Yu, H., & Du, S. (2021). Identification and characterization of key haem pathway genes associated with the synthesis of porphyrin in Pacific oyster (*Crassostrea gigas*). *Comparative Biochemistry and Physiology Part - B: Biochemistry and Molecular Biology, 255(March)*. https://doi.org/10.1016/j.cbpb.2021.110595

Jablonski, N. G., & Chaplin, G. (2010). Human skin pigmentation as an adaptation to UV radiation. *Proceedings of the National Academy of Sciences, 107(Supplement_2),* 8962–8968. https://doi.org/10.1073/pnas.0914628107

Jiang, K., Jiang, L., Nie, H., & Huo, Z. (2020). Molecular cloning and expression analysis of tyrosinases (tyr) in four shell-color strains of Manila clam *Ruditapes philippinarum. PeerJ, 8,* 1–20. https://doi.org/10.7717/peerj.8641

Klopfenstein, D. V., Zhang, L., Pedersen, B. S., Ramírez, F., Vesztrocy, A. W., Naldi, A., Mungall, C. J., Yunes, J. M., Botvinnik, O., Weigel, M., Dampier, W., Dessimoz, C., Flick, P., & Tang, H. (2018). GOATOOLS: A Python library for Gene Ontology analyses. *Scientific Reports, 8(1)*, 1–17. https://doi.org/10.1038/s41598-018-28948-z

Kobayashi, T., Kawahara, I., Hasekura, O., & Kijima, A. (2004). Genetic control of bluish shell color variation in the Pacific abalone, *Haliotis discus hannai. Journal of Shellfish Research, 23(4),* 1153–1157.

Ky, C. L., Blay, C., Sham-Koua, M., Vanaa, V., Lo, C., & Cabral, P. (2013). Family effect on cultured pearl quality in black-lipped pearl oyster *Pinctada margaritifera* and insights for genetic improvement. *Aquatic Living Resources, 26,* 133–145. https://doi.org/10.1051/alr/2013055

Ky, C. L., Blay, C., Sham-Koua, M., Lo, C., & Cabral, P. (2014). Indirect improvement of pearl grade and shape in farmed *Pinctada margaritifera* by donor "oyster" selection for green pearls. *Aquaculture, 432(30),* 154–162. https://doi.org/10.1016/j.aquaculture.2014.05.002

Ky, C. L., Demmer, J., Blay, C., & Lo, C. (2015). Age-dependence of cultured pearl grade and colour in the black-lipped pearl oyster *Pinctada margaritifera. Aquaculture Research, 48,* 955–968. https://doi.org/10.1111/are.12938

Ky, C. L., Nakasai, S., Pommier, S., Sham Koua, M., & Devaux, D. (2016). The Mendelian inheritance of rare flesh and shell colour variants in the black-lipped pearl oyster (*Pinctada margaritifera*). *Animal Genetics, 47(5),* 610–614. https://doi.org/10.1111/age.12454

Ky, C. L., Le Pabic, L., Sham Koua, M., Nicolas, M., Seiji, N., & Devaux, D. (2017a). Is pearl colour produced from *Pinctada margaritifera* predictable through shell phenotypes and rearing environments selections ? *Aquaculture, 48(3),* 1041–1057. https://doi.org/https://doi.org/10.1111/are.12947

Ky, C. L., Lo, C., & Planes, S. (2017b). Mono- and polychromatic inner shell phenotype diversity in *Pinctada margaritifera* donor pearl oysters and its relation with cultured pearl colour. *Aquaculture, 468,* 199–205. https://doi.org/10.1016/j.aquaculture.2016.10.017

Ky, C. L., Quillien, V., Broustal, F., Soyez, C., & Devaux, D. (2018). Phenome of pearl quality traits in the mollusc transplant model *Pinctada margaritifera. Scientific Reports, 8(1),* 1–11. https://doi.org/10.1038/s41598-018-20564-1

Lado, J., Cronje, P., Alquézar, B., Page, A., Manzi, M., Gómez-Cadenas, A., Stead, A. D., Zacarías, L., & Rodrigo, M. J. (2015). Fruit shading enhances peel color, carotenes accumulation and chromoplast differentiation in red grapefruit. *Physiologia Plantarum, 154(4),* 469–484. https://doi.org/10.1111/ppl.12332

Latchere, O., Mehn, V., Gaertner-Mazouni, N., Le Moullac, G., Fievet, J., Belliard, C., Cabral, P., Saulnier, D. (2018). Influence of water temperature and food on the last stages of cultured pearl mineralization from the black-lip pearl oyster *Pinctada margaritifera*. *PloS One, 13(3),* e0193863. https://doi.org/10.1371/journal.pone.0193863

Lemer, S., Saulnier, D., Gueguen, Y., & Planes, S. (2015). Identification of genes associated with shell color in the black-lipped pearl oyster, *Pinctada margaritifera. BMC Genomics, 16(1)*, 568. https://doi.org/10.1186/s12864-015-1776-x

Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60. [PMID: 19451168]

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, Z., Li, Q., Liu, S., Han, Z., Kong, L., & Yu, H. (2021). Integrated Analysis of Coding Genes and Non - coding RNAs Associated with Shell Color in the Pacific Oyster (*Crassostrea gigas*). *Marine Biotechnology*, (0123456789), 1–13. https://doi.org/10.1007/s10126-021-10034-7

Lichtenthaler, H. K. (1987). Chlorophylls and carotenoids: pigments of photosynthetic biomembranes. *Methods in Enzymology, 148(C),* 350–382. https://doi.org/10.1016/0076-6879(87)48036-1

Lietz, G., Oxley, A., Leung, W., & Hesketh, J. (2012). Single Nucleotide Polymorphisms Upstream Gene Influence Provitamin A Conversion Efficiency in Female Volunteers 1 – 4. *The Journal of Nutrition,* 12–16. https://doi.org/10.3945/jn.111.140756

Liu, F., Visser, M., Duffy, D. L., Hysi, P. G., Jacobs, L. C., Lao, O., Zhong, K., Walsh, S., Chaitanya, L., Wollstein, A., Zhu, G., Montgomery, G. W., Henders, A. K., Mangino, M., Glass, D., Bataille, V., Sturm, R. A., Rivadeneira, F., Hofman, A., van IJcken, W. F. J., Uitterlinden, A. G., Palstra R. T. S., Spector, T. D., Martin, N. G., Nijsten, T. E. C., & Kayser, M. (2015). Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. *Human Genetics, 134(8),* 823–835. https://doi.org/10.1007/s00439-015-1559-0

Lyons, L. A., Foe, I. T., Rah, H. C., & Grahn, R. A. (2005). Chocolate coated cats: TYRP1 mutations for brown color in domestic cats. *Mammalian Genome, 16(5),* 356–366. https://doi.org/10.1007/s00335-004-2455-4

Ma, Z., Huang, J., Sun, J., Wang, G., Li, C., Xie, L., & Zhang, R. (2007). A novel extrapallial fluid protein controls the morphology of nacre lamellae in the pearl oyster, *Pinctada fucata. Journal of Biological Chemistry, 282(32),* 23253–23263. https://doi.org/10.1074/jbc.M700001200

Marie, B., Joubert, C., Belliard, C., Tayale, A., Zanella-Cléon, I., Marin, F., Gueguen, Y., & Montagnani, C. (2012). Characterization of MRNP34, a novel methionine-rich nacre protein from the pearl oysters. *Amino Acids, 42(5),* 2009–2017. https://doi.org/10.1007/s00726-011-0932-0

McCormick, R. F., Truong, S. K., & Mullet, J. E. (2015). RIG: Recalibration and interrelation of genomic sequence data with the GATK. *G3: Genes, Genomes, Genetics, 5(4),* 655–665. https://doi.org/10.1534/g3.115.017012

Meyer, U. A., Strand, L. J., Doss, M., Rees, A. C., & Marver, H. S. (1972). Intermittent acute porphyria—demonstration of a genetic defect in porphobilinogen metabolism. *New England Journal of Medicine, 286(24),* 1277–1282. https://doi.org/10.1056/NEJM197206152862401

Müller, H., Jimenez-Heredia, R., Krolo, A., Hirschmugl, T., Dmytrus, J., Boztug, K., & Bock, C. (2017). VCF.Filter: Interactive prioritization of disease-linked genetic variants from sequencing data. *Nucleic Acids Research, 45(W1),* W567–W572. https://doi.org/10.1093/nar/gkx425

Nadin, L., & Murray, M. (1999). Participation of CYP2C8 in retinoic acid 4-hydroxylation in human hepatic microsomes. *Biochemical Pharmacology, 58(7),* 1201–1208. https://doi.org/10.1016/S0006-2952(99)00192-6

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., & Durbin, R. (2016). BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics, 32(11),* 1749–1751. https://doi.org/10.1093/bioinformatics/btw044

Naveed, A., Li, H., & Liu, X. (2018). Cytochrome P450s: Blueprints for Potential Applications in Plants. *Journal of Plant Biochemistry & Physiology, 06(01),* 1–9. https://doi.org/10.4172/2329-9029.1000204

Ng, A., Uribe, R. A., Yieh, L., Nuckels, R., & Gross, J. M. (2009). Zebrafish mutations in gart and paics identify crucial roles for de novo purine synthesis in vertebrate pigmentation and ocular development. *Development, 136(15),* 2601–2611. https://doi.org/10.1242/dev.038315

Nie, H., Jiang, K., Jiang, L., Huo, Z., Ding, J., & Yan, X. (2020). Transcriptome analysis reveals the pigmentation related genes in four different shell color strains of the Manila clam *Ruditapes philippinarum*. *Genomics, 112(2),* 2011–2020. https://doi.org/10.1016/j.ygeno.2019.11.013

Pante E, Simon-Bouhet B (2013). "marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R." *PLoS ONE, 8(9),* e73051. doi:10.1371/journal.pone.0073051

Pante, E., & Simon-bouhet, B. (2019). Package ' marmap ' : Import, Plot and Analyze Bathymetric and Topographic Data. CRAN.

Pimsler, M. L., Jackson, J. M., & Lozier, J. D. (2017). Population genomics reveals a candidate gene involved in bumble bee pigmentation. *Ecology and evolution, 7(10),* 3406–3413. https://doi.org/10.1002/ece3.2935

Phillips, N. (2017). Package ' yarrr ': YaRrr!: The Pirate's Guide to R. https://bookdown.org/ndphillips/YaRrr

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics, 26(6),* 841–842. https://doi.org/10.1093/bioinformatics/btq033

Reisser, C. (2018). Pooled samples of mantle tissues from *P. margaritifera* – Project AMELIGEN color Fev. 2018. https://doi.org/https://sextant.ifremer.fr/record/e0a8f7bc-423d-4043-b997-d839a083f01b

Reisser, C.M.O, Vidal-Dupiol, J., Ky, C.L., Stenger, P.L. (2020). *Pinctada margaritifera* reference genome V2. IFREMER, https://doi.org/10.12770/d19f9fc1-eb67-4419-88a4c0586a73e25c

Céline Reisser, Jérémie Vidal-Dupiol, Chin Long-Ky, Pierre-Louis Stenger (2020). *Pinctada margaritifera* reference genome V2. IFREMER https://doi.org/10.12770/d19f9fc1-eb67-4419-88a4-c0586a73e25c

Céline Reisser, Jérémie Vidal-Dupiol, Chin Long-Ky, Pierre-Louis Stenger (2020). *Pinctada margaritifera* reference genome V2. IFREMER https://doi.org/10.12770/d19f9fc1-eb67-4419-88a4-c0586a73e25c

Richards, P. M., Liu, M. M., Lowe, N., Davey, J. W., Blaxter, M. L., & Davison, A. (2013). RADSeq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Molecular Ecology, 22(11)*, 3077–3089. https://doi.org/10.1111/mec.12262

Rousseau, M., & Rollion-Bard, C. (2012). Influence of the depth on the shape and thickness of nacre tablets of *Pinctada margaritifera* pearl oyster, and on oxygen isotopic composition. *Minerals, 2(4),* 55–64. https://doi.org/10.3390/min2010055

Shamim, G., Ranjan, S. K., Pandey, D. M., & Ramani, R. (2014). Biochemistry and biosynthesis of insect pigments. *European Journal of Entomology*, 111, 149–164. https://doi.org/10.14411/eje.2014.021

Schneider-Yin, X., Hergersberg, M., Goldgar, D. E., Rüfenacht, U. B., Schuurmans, M. M., Puy, H., Deybach, J. C., Minder, E. I. (2002). Ancestral founder of mutation W283X in the porphobilinogen deaminase gene among acute intermittent porphyria patients. *Human Heredity, 54(2),* 69–81. https://doi.org/10.1159/000067665

Sculley, D. G., Dawson, P. A., Emmerson, B. T., & Gordon, R. B. (1992). A review of the molecular basis of hypoxanthine-guanine phosphoribosyltransferase (HPRT) deficiency. *Human Genetics, 90(3),* 195–207. https://doi.org/10.1007/BF00220062

Searle, S. R., & Gruber, M. H. (1971). Linear models (Wiley). New York, USA.

Shi, Y., Yu, C., Gu, Z., Zhan, X., Wang, Y., & Wang, A. (2013). Characterization of the pearl oyster (*Pinctada martensii*) mantle transcriptome unravels biomineralization genes. *Marine Biotechnology, 15(2),* 175–187. https://doi.org/10.1007/s10126-012-9476-x

Shima, Y., Teruya, K., & Ohta, H. (2006). Association between intronic SNP in urate-anion exchanger gene, SLC22A12, and serum uric acid levels in Japanese. *Life Sciences, 79(23),* 2234–2237. https://doi.org/10.1016/j.lfs.2006.07.030

Shin, J. (2012). Immunohistochemical Evaluation and Gene Expression Profiling of Senile Lentigo. Master's thesis, Graduate School of Medicine, Ajou University, Korea. http://repository.ajou.ac.kr/handle/201003/7582

Shinohara, M., Kinoshita, S., Tang, E., Funabara, D., & Kakinuma, M. (2018). Comparison of two pearl sacs formed in the same recipient oyster with different genetic background involved in yellow pigmentation in *Pinctada fucata*. *Marine Biotechnology, 20(5),* 594– 602.

Siersema, P. D., Roolj, F. W. U. De, Edbthoven-bosdijk, A., & Wilson, J. H. P. (1990). Erythrocyte porphobilinogen deaminase activity in porphyria cutanea tarda. *Clinical Chemistry, 36(10)*, 1779–1783.

Slatkin, M. (1978). Spatial patterns in the distributions of polygenic characters. *Journal of Theoretical Biology, 70(2),* 213–228. https://doi.org/10.1016/0022-5193(78)90348-X

Sokolov, E. P. (2000). An improved method for DNA isolation from mucopolysaccharide rich molluscan tissues. *Journal of Molluscan Studies, 66(4),* 573–575. https://doi.org/10.1093/mollus/66.4.573

Song, G., Li, Y., Cheng, C., Zhao, Y., Gao, A., Zhang, R., … Liu, Z. (2009). Structural insight into acute intermittent porphyria. *The FASEB Journal, 23(2),* 396–404. https://doi.org/10.1096/fj.08-115469

Song, J., Li, Q., Zhong, X., Kong, L., & Yu, H. (2017). Genetic diversity and outlier loci detecting of shell color variation in the Pacific oyster (*Crassostrea gigas*) by SNP markers. *Aquatic Living Resources, 30,* 10. https://doi.org/10.1051/alr/2017009

Sonthalia, S., Daulatabad, D., & Sarkar, R. (2016). Glutathione as a skin whitening agent : Facts, myths, evidence and controversies. *Indian Journal of Dermatology, 82(3),* 262. https://doi.org/10.4103/0378-6323.179088

Srovnalova, A., Svecarova, M., Kopecna Zapletalova, M., Anzenbacher, P., Bachleda, P., Anzenbacherova, E., & Dvorak, Z. (2014). Effects of anthocyanidins and anthocyanins on the expression and catalytic activities of CYP2A6, CYP2B6, CYP2C9, and CYP3A4 in primary human hepatocytes and human liver microsomes. *Journal of Agricultural and Food Chemistry, 62(3),* 789–797. https://doi.org/10.1021/jf404643w

Stahl, E. A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B. F., Kraft P., Chen, R., Kallberg, H. J., Kurreeman, F. A. S., Diabetes Genetics Replication and Metaanalysis Consortium, Myocardial Infarction Genetics Consortium, Kathiresan, S., Wijmenga, C., Gregersen, P. K., Alfredsson, L., Siminovitch, K. A., Worthington, J., de Bakker, P. I. W., Raychaudhuri, S., & Plenge, R. M. (2012). Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature Genetics, 44(5),* 483–489. https://doi.org/10.1038/ng.2232

Stenger, P. L. (2017). Package ImaginR. CRAN. Retrieved from https://cran.rproject.org/web/packages/ImaginR/ImaginR.pdf

Stenger, P. L., Vidal-Dupiol, J., Reisser, C., Planes, S., & Ky, C. (2019). Colour plasticity in the shells and pearls of animal graft model *Pinctada margaritifera* through colour quantification with the HSV system. *Scientific Reports, (9),* 75:20. https://doi.org/https://doi.org/10.1038/s41598-019-43777-4

Stenger, P. L., Ky, C. L., Reisser, C., Duboisset, J., Dicko, H., Durand, P., Quintric, L., Planes, S., Vidal-Dupiol, J. (2021a). Molecular pathways and pigments underlying the colors of the pearl oyster *Pinctada margaritifera var. cumingii* (Linnaeus 1758). *Genes, 12(3),* 421. https://doi.org/10.3390/genes12030421

Stenger, P. L., Ky, C. L., Reisser, C. M. O., Cosseau, C., Grunau, C., Mege, M., Planes, S., VidalDupiol, J. (2021b). Environmentally driven color variation in the pearl oyster *Pinctada margaritifera var. cumingii* (Linnaeus, 1758) is associated with differential methylation of CpGs in pigment- and biomineralization-related genes. *Frontiers in Genetics, 12(March),* 1–18. https://doi.org/10.3389/fgene.2021.630290

Sturm, R. A., Duffy, D. L., Zhao, Z. Z., Leite, F. P. N., Stark, M. S., Hayward, N. K. K., Martin, N. G., & Montgomery, G. W. (2008). A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *American Journal of Human Genetics, 82(2),* 424–431. https://doi.org/10.1016/j.ajhg.2007.11.005

Sun, X., Yang, A., Wu, B., Zhou, L., & Liu, Z. (2015). Characterization of the mantle transcriptome of Yesso scallop (*Patinopecten yessoensis*): Identification of genes potentially involved in biomineralization and pigmentation. *PLoS ONE, 10(4),* 1–19. https://doi.org/10.1371/journal.pone.0122967

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE, 6(7).* https://doi.org/10.1371/journal.pone.0021800

Tan, A., Abecasis, G. R., & Kang, H. M. (2015). Unified representation of genetic variants. *Bioinformatics, 31(13),* 2202–2204. https://doi.org/10.1093/bioinformatics/btv112

Theil, E. C., Chen, H., Miranda, C., Janser, H., Elsenhans, B., Núñez, M. T., … Schümann, K. (2012). Absorption of iron from ferritin is independent of heme iron and ferrous salts

in women and rat intestinal segments. *The Journal of Nutrition, 142(3),* 478–483. https://doi.org/10.3945/jn.111.145854

Tokuhiro, S., Yamada, R., Chang, X., Suzuki, A., Kochi, Y., Sawada, T., Suzuki, M., Nagasaki, M., Ohtsuki, M., Ono, M., Furukawa, H., Nagashima, M., Yoshino, S., Mabuchi, A., Sekine, A., Saito, S., Takahashi, A., Tsunoda, T., Nakamura Y., & Yamamoto, K. (2003). An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nature Genetics, 35(4),* 341–348. https://doi.org/10.1038/ng1267

Tolmach, J., & Graham, N. (1942). Pigmentation of the skin due to vitamin A deficiency. *Glasgow Medical Journal, 30(8),* 288–293.

Wacewicz, M., Socha, K., Soroczyńska, J., Niczyporuk, M., Aleksiejczuk, P., Ostrowska, J., & Borawska, M. H. (2018). Selenium, zinc, copper, Cu/Zn ratio and total antioxidant status in the serum of vitiligo patients treated by narrow-band ultraviolet-B phototherapy. *Journal of Dermatological Treatment, 29(2),* 190–195. https://doi.org/10.1080/09546634.2017.1357797

Wang, A., Wang, Y., Gu, Z., Li, S., Shi, Y., & Guo, X. (2011). Development of expressed sequence tags from the pearl oyster, *Pinctada martensii,* Dunker. *Marine Biotechnology, 13(2),* 275–283. https://doi.org/10.1007/s10126-010-9296-9

Watt, W. B. (1972). Xanthine dehydrogenase and pteridine metabolism in *Colias* butterflies. *Journal of Biological Chemistry, 247,* 1445–1451. https://doi.org/10.1016/S0021-9258(19)45578-6

Weston, & Calaway. (2019). Package ' doParallel ' : Foreach Parallel Adaptor for the "parallel" Package. https://cran.r-project.org/web/packages/doParallel/index.html

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org

Wickham, H. (2019). stringr: Simple, Consistent Wrappers for Common String Operations. Retrieved from https://cran.r-project.org/web/packages/stringr/stringr.pdf

Wickham H, Girlich M (2022). tidyr: Tidy Messy Data. https://tidyr.tidyverse.org, https://github.com/tidyverse/tidyr

Wickham H, Hester J, Bryan J (2022a). readr: Read Rectangular Text Data. https://readr.tidyverse.org, https://github.com/tidyverse/readr

Wickham H, François R, Henry L, Müller K (2022b). dplyr: A Grammar of Data Manipulation. https://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr

Williams, S. T. (2016). Molluscan shell colour. *Biological Reviews.* https://doi.org/10.1111/brv.12268

Winkler, F. M., Estévez, B. F., Jollán, L. B., & Garrido, J. P. (2001). Inheritance of the general shell color in the scallop *Argopecten purpuratus* (Bivalvia*: Pectinidae*). *Journal of Heredity, 92(6),* 521–525. https://doi.org/10.1093/jhered/92.6.521

Xu, M., Huang, J., Shi, Y., Zhang, H., & He, M. (2019). Comparative transcriptomic and proteomic analysis of yellow shell and black shell pearl oysters, *Pinctada fucata martensii*. *BMC Genomics, 20(1),* 1–14. https://doi.org/10.1186/s12864-019-5807-x

Yano, M., Nagai, K., Morimoto, K., & Miyamoto, H. (2006). Shematrin: A family of glycinerich structural proteins in the shell of the pearl oyster *Pinctada fucata. Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology, 144(2),* 254–262. https://doi.org/10.1016/j.cbpb.2006.03.004

Yuan, C., Mao, J., Sun, H., Wang, Y., Guo, M., Wang, X., … Chang, Y. (2021). Genome-wide DNA methylation profile changes associated with shell coloration in the Yesso scallop (*Patinopecten yessoensis*) as measured by whole-genome bisulfite sequencing. *BMC Genomics, 22(1),* 1–14. https://doi.org/10.1186/s12864-021-08055-6

Yue, X., Nie, Q., Xiao, G., & Liu, B. (2015). Transcriptome analysis of shell color-related genes in the clam *Meretrix meretrix. Marine Biotechnology, 17(3),* 364–374. https://doi.org/10.1007/s10126-015-9625-0

Zheng, H., Zhang, T., Sun, Z., Liu, W., & Liu, H. (2013). Inheritance of shell colours in the noble scallop *Chlamys nobilis* (Bivalve: *Pectinidae*). *Aquaculture Research, 44(8),* 1229–1235. https://doi.org/10.1111/j.1365-2109.2012.03124.x

Figures & Tables

Table 1:

Number of sequences after the different bioinformatic steps for each of the twelve pools and the number of sequences that are mapped against the reference genome.

The last row gives the mean of each column.

Color phenotype	Geographical site	Raw sequences	After trimming	Percent of deleted sequences	Number of properly paired against the genome	Percent of properly paired against the genome	Number of properly paired after filters	Percent of properly paired after filters
Red	**Takapoto**	933451922	804525582	16.02	652910874	81.15	491651124	75.30
Red	**Katiu**	919931270	825237588	11.47	675882634	81.90	506143676	74.89
Red	**Gambier**	977173146	900905220	8.46	746316126	82.84	545397822	73.08
Red	**Hatchery**	875890480	757998130	15.55	631284132	83.28	472079359	74.78
Yellow	**Takapoto**	938782098	821910616	14.22	671154850	81.66	507808553	75.66
Yellow	**Katiu**	943199888	866874884	8.80	709770466	81.88	522722833	73.65
Yellow	**Gambier**	978499586	884295558	10.65	729092800	82.45	540454292	74.13
Yellow	**Hatchery**	898214928	790138682	13.68	653342130	82.69	490391020	75.06
Green	**Takapoto**	911432464	785142178	16.08	632289464	80.53	472054295	74.66
Green	**Katiu**	898068018	785537094	14.32	649464162	82.68	475609233	73.23
Green	**Gambier**	949968516	867379084	9.52	713228138	82.23	530586280	74.39
Green	**Hatchery**	901035306	789581510	14.11	654043378	82.83	490202797	74.95
	Mean	927137301	823293843	12.74	676564929	82.18	503758440	74.48

Table 2:

Enriched GO terms in color phenotypes according to GOATOOLS analyses, listing the terms found in the three phenotypes, those found in two phenotypes and those that are phenotype specific.

Enriched GO terms in common for the three phenotypes
protein binding (GO:0005515)
binding (GO:0005488)
catalytic activity (GO:0003824)
ion binding (GO:0043167)
anion binding (GO:0043168)
small molecule binding (GO:0036094)
nucleotide binding (GO:0000166)
ribonucleotide binding (GO:0032553)
purine nucleotide binding (GO:0017076)
purine ribonucleoside triphosphate binding (GO:0035639)
nucleoside phosphate binding (GO:1901265)
enzyme binding (GO:0019899).
Enriched GO terms in common between red and green phenotype
purine ribonucleotide binding (GO:0032550)
identical protein binding (GO:0042802)
transcription regulatory region sequence-specific DNA binding (GO:0000976)
Enriched GO terms in common between red and yellow phenotype
drug binding (GO:0008144)
carbohydrate derivative binding (GO:0097367)
catalytic activity(GO:0003824)
Enriched GO terms in common between green and yellow phenotype
transporter activity (GO:0005215)
transmembrane activity (GO:0022857)
ion transmembrane transporter activity (GO:0015075)
inorganic molecular entity transmembrane transporter activity (GO:0015318)
Enriched GO terms red specific adenyl nucleotide binding (GO:0030554)
adenyl ribonucleotide binding (GO:0032559)
ATP binding (GO:0005524)
hydroxymethylbilane synthase activity (GO:0004418)
protein dimerization activity (GO:0046983)
protein heterodimerization activity (GO:0046982)
DNA-binding transcription factor activity (GO:0003700)
enzyme regulator activity (GO:0030234)
anion transmembrane transporter activity (GO:0008509)
transcription regulator activity (GO:0140110)
regulatory region nucleic acid binding (GO:0001067)
Enriched GO terms yellow specific calcium ion binding (GO:0005509)
G protein-coupled receptor activity (GO:0004930)
molecular transducer activity (GO:0060089)
cation binding (GO:0043169)
metal ion binding (GO:0046872)
signaling receptor activity (GO:0038023)
transmembrane signaling receptor activity (GO:0004888)
transferase activity (GO:0016740)
hydrolase activity (GO:0016787)
phosphoric diester hydrolase activity (GO:0008081)
hydrolase activity acting on ester bonds (GO:0016788)

Enriched GO terms green specific organic cyclic compound binding (GO:0097159)
heterocyclic compound binding (GO:1901363)
coenzyme binding (GO:0050662)
cytoskeletal protein binding (GO:0008092)
oxidoreductase activity (GO:0016491)
sequence-specific DNA binding (GO:0043565)
double-stranded DNA binding (GO:0003690)
cofactor binding (GO:0018160)
RNA polymerase II regulatory region DNA binding (GO:0001012)

Table 3:

Name of pigment related proteins impacted by SNPs specific to the green, yellow and green phenotype. For each protein, the information for one SNP is given as an example, with the name of this SNP (scaffold number and position of the SNP), the reference ® and alternative (A) allele, their impact, and the number of transcripts bearing SNPs coding the same protein.

Gene name	SNPs number		SNP	R	A	Impact	Number of other similar genes
Red specific							
Beta,beta-carotene 15,15'-dioxygenase	1	Example for one SNP	scaffold605|size173110_50152	A	T	Upstream gene	0
Bifunctional purine biosynthesis protein PURH	1		scaffold8173|size41510_24501	T	G	Upstream gene	0
Cytochrome P450 26A1-like	1		scaffold6003|size54905_33122	A	C	Intergenic	0
Cytochrome P450 2C8	4		scaffold417|size388363_270987	A	T	Intergenic	1
Cytochrome P450 3A29	1		scaffold2294|size169326_22277	C	T	missense_variant	0
Cytochrome P450 3A4	1		scaffold1969|size103307_48966	T	C	Intron	0
Cytochrome P450 4F8-like	1		scaffold655|size199117_114782	C	A	Upstream gene	0
Porphobilinogen deaminase	14		scaffold2460|size144317_61211	G	A	Upstream gene	4 (see below)
Porphobilinogen deaminase	4		scaffold2460|size144317_128208	T	C	Intergenic	4 (see above & below)
Porphobilinogen deaminase	2		scaffold5737|size147992_39135	A	C	Downstream gene	4 (see above & below)
Porphobilinogen deaminase-like	1		scaffold1000|size145372_49694	C	A	Intergenic	4 (see above & below)
Porphobilinogen deaminase-like	3		scaffold2460|size144317_56369	T	C	Upstream gene	4 (see above)
Shematrin-like protein 1	13		scaffold2460|size144317_86009	T	A	Upstream gene	0
Tyrosinase 2	1		scaffold1567|size137386_55806	G	A	Upstream gene	0
UDP-glucuronosyltransferase 1-2	1		scaffold953|size134644_10904	T	A	Downstream gene	0
Versicolorin reductase-like	1		scaffold4800|size171772_160414	C	T	Intergenic	0
Yellow specific							
Amorphous calcium carbonate binding protein 1	4		scaffold97|size234723_13738	T	A	Upstream gene	1
Cadherin-23 isoform X1	9		scaffold5945|size55440_53306	C	G	Upstream gene	4
Copper/zinc superoxide dismutase	47		scaffold1412|size117757_38094	C	T	Upstream gene	37
Cytochrome b5 reductase 4-like isoform X3	1	SNP	scaffold3263|size165871_73814	T	A	Intergenic	0

Cytochrome P450 2C8-like	1	Example for one	scaffold417|size388363_126041	T	G	Upstream gene	0
Cytochrome P450 3A11-like isoform X1	2		scaffold9123|size36889_30428	G	A	Intron	1
Cytochrome P450 4F8-like	1		scaffold5925|size155237_135045	C	T	Intergenic	0
Hypoxanthine-guanine phosphoribosyltransferase	2		scaffold1749|size108746_82049	T	C	Intron	0
Laccase-4-like isoform X1	1		scaffold337|size296661_26903	T	C	Upstream gene	0
Laccase-6	1		scaffold4044|size72409_7801	T	G	Upstream gene	0
Tyrosinase B4	2		scaffold2426|size145584_113288	C	T	Intergenic	1
Tyrosinase BPmax1	2		scaffold5775|size102228_21643	G	A	Intergenic	0
Tyrosinase-like protein 1	4		scaffold1797|size279199_75418	A	G	Intergenic	1
Tyrosine-protein kinase PR2 isoform X1	2		scaffold10826|size62665_1539	A	G	Downstream gene	0
Tyrosine-protein kinase transmembrane receptor Ror	1		scaffold1018|size159565_123029	A	T	Intron	0
Tyrosine-protein phosphatase non-receptor type 11	1		scaffold5286|size98312_75622	T	A	Intron	0
Tyrosine-protein phosphatase non-receptor type 23	2		scaffold4726|size85232_2882	C	T	Intergenic	0
Visual pigment-like receptor peropsin	1		scaffold1490|size115539_64221	T	A	Intergenic	0
Vitamin D(3) 25-hydroxylase	1		scaffold5047|size126357_6322	C	A	Upstream gene	0
Xanthine dehydrogenase/oxidase	1		scaffold5976|size55229_32252	A	G	Downstream gene	0
Green specific							
Cadherin-23	4	Example for one SNP	scaffold910|size252188_14555	T	C	Intron	3
Calcium-binding protein CML8	1		scaffold88|size277366_192751	A	G	Intron	0
Calcium-binding protein LPS1-alpha-like	1		scaffold14161|size18149_12593	A	T	Intron	0
Cytochrome P450 20A1	2		scaffold791|size153329_87835	T	C	Intron	0
Cytochrome P450 2C8-like	4		scaffold49|size265340_38705	A	C	Intron	3
Cytochrome P450 4A25	1		scaffold1341|size119781_6168	C	T	Downstream gene	0

Cytochrome P450 4F12-like	2		scaffold1676|size364098_330215	A	T	Intergenic	0
Ferritin 1	1		scaffold551|size158850_20422	A	C	Upstream gene	0
Ferrochelatase, mitochondrial	1		scaffold5249|size131904_104547	G	A	Intergenic	0
Flavin reductase (NADPH)	1		scaffold2158|size160354_118513	T	C	Intergenic	0
Glutathione S-transferase 1-like	1		scaffold1115|size353168_63863	A	T	Intron	0
Perlucin-like protein	4		scaffold3547|size129194_105367	A	T	Downstream gene	2
Tyrosinase 1	2		scaffold454|size172198_160132	T	C	Intergenic	0
Tyrosinase B2.1	1		scaffold1980|size291657_59414	C	G	Intron	0
Urease subunit alpha-like	2		scaffold101|size233448_106064	A	G	Intron	0
Visual pigment-like receptor peropsin	1		scaffold2909|size388074_257210	T	A	Downstream gene	0

Figure 1

Map of French Polynesia showing the sampling sites and number of sequenced samples (N) per phenotype in this study: The hatchery individuals originated from the reproduction of wild individuals from the Gambier archipelago but were raised in the Ifremer lagoon concession in Tahiti.

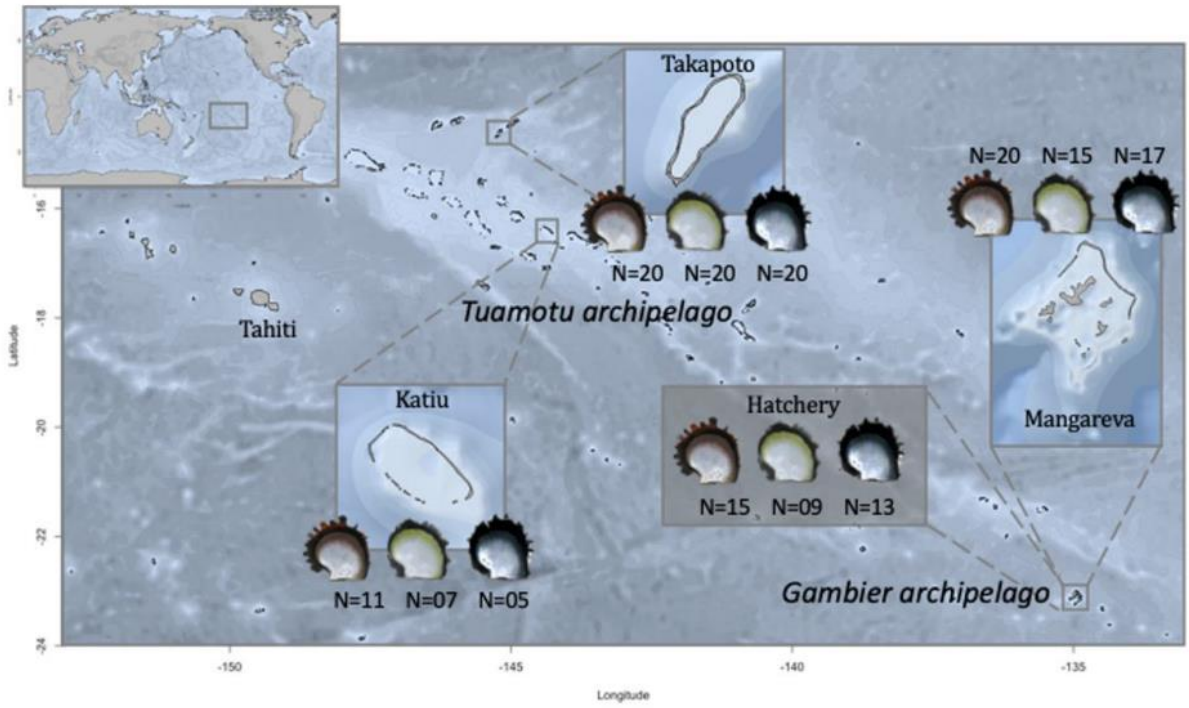


Figure 2

Principal Component Analysis (PCA) of the twelve pooled samples based on Allelic Frequencies (AF) at the 22,171,628 SNPs loci

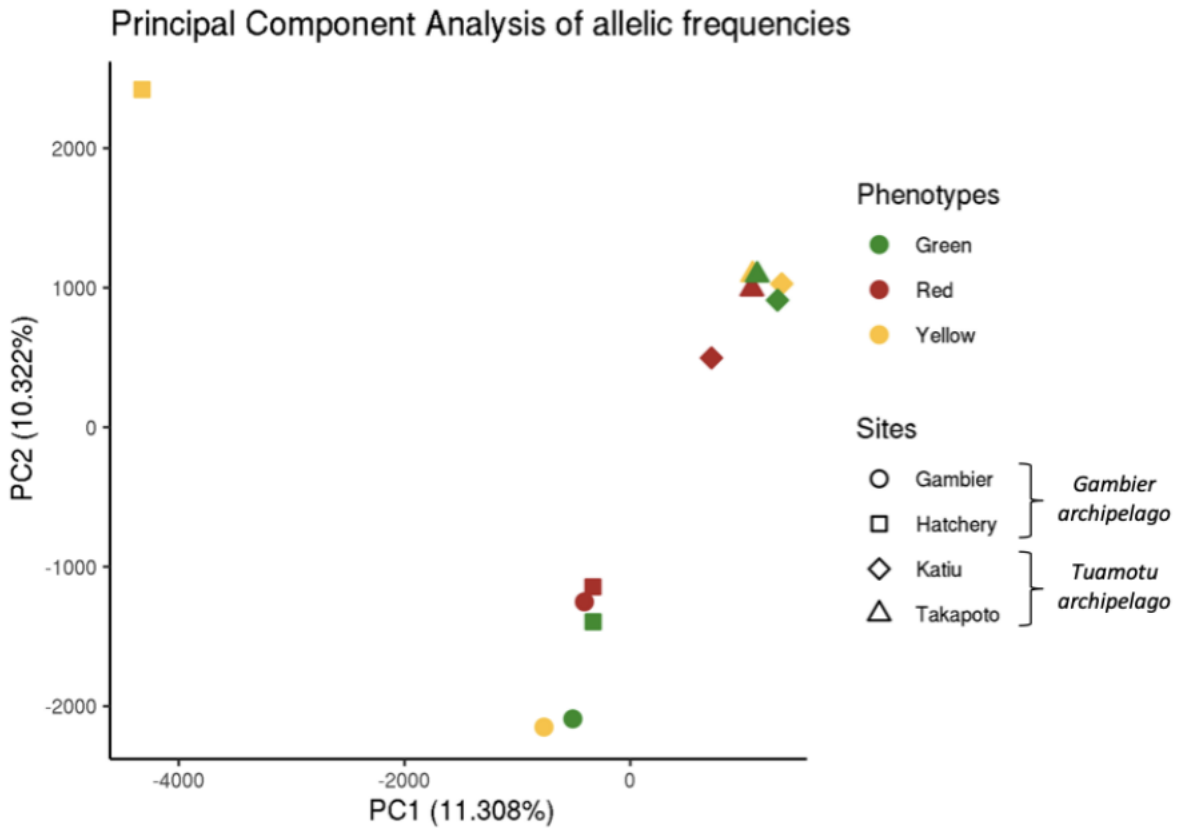


Figure 3

Chord diagram of significant SNPs (P value < 1e10-9) obtained across the geographic locations and color phenotypes. Grey chords represent the SNPs associated to geography. The color associated SNPs are indicated by green chords, while color-only associated SNPs (SNPs that were not also geographically associated) are delimited by black borders.

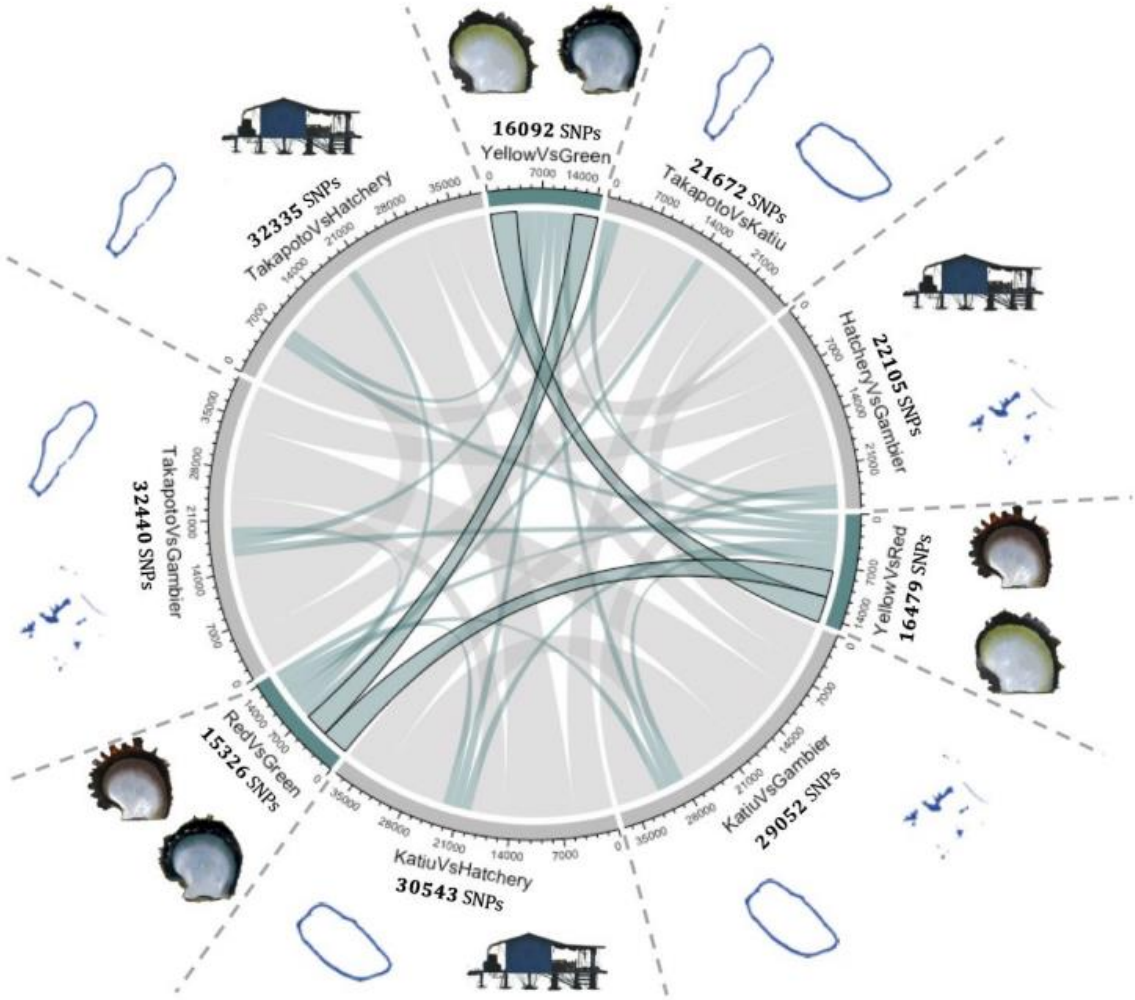


Figure 4

Allelic frequencies distributions of the color-only associated SNPs according to their original pool. Histograms shows the AF distribution with 30 bins, while pirate plots (raw data, descriptive statistics, and inferential statistics plotted in same a graphics) show raw AF distribution data (points), the median (bar/line), the 95% HDI (High-Dimensional Inference) (moustache) and the data density. Letters beside these plots correspond to the significance in the difference by pairwise Wilcoxon test.

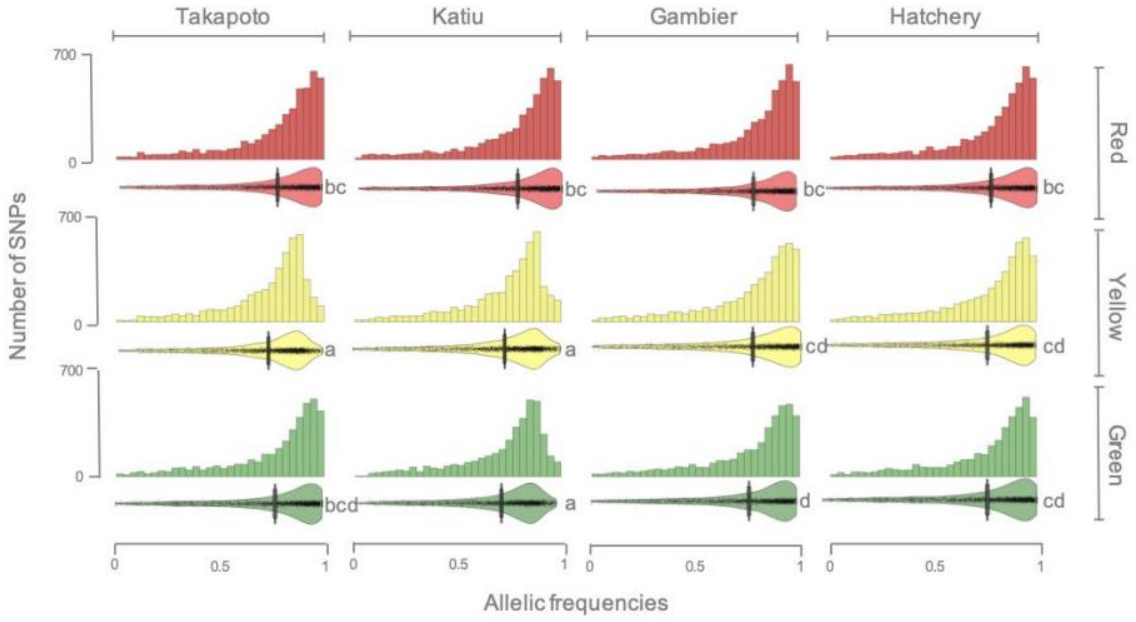


Figure 5

Number of SNPs impacts depending on pairwise color phenotype comparisons. The total number of effects contains all shown impact categories of the plot as well as low, missense, nonsense and silent impacts.

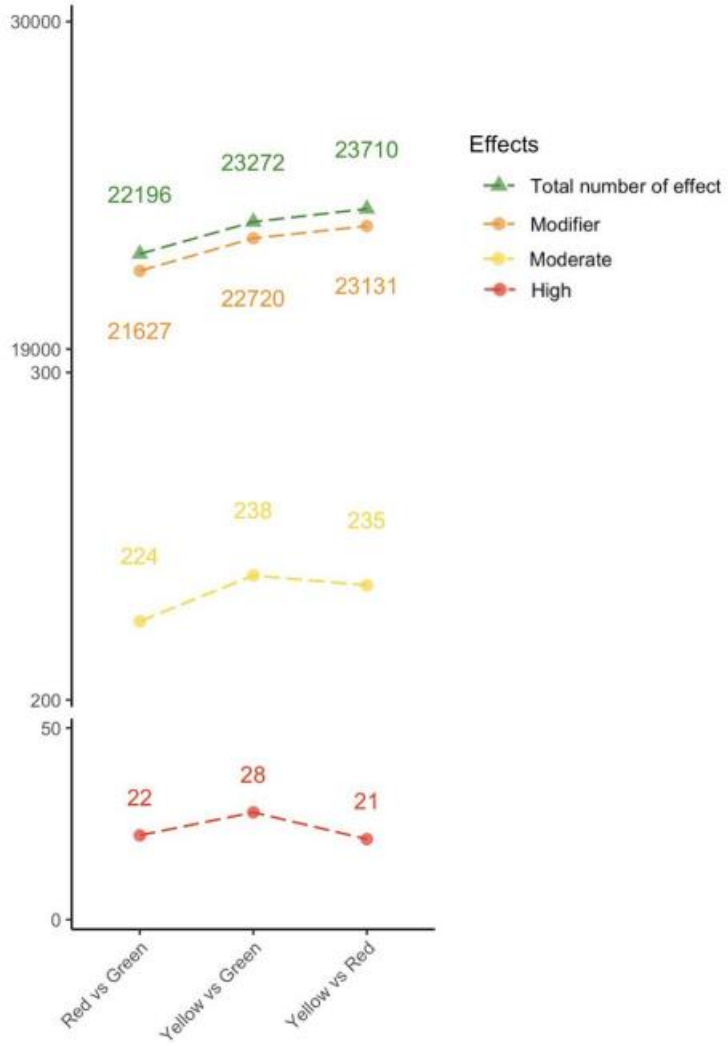
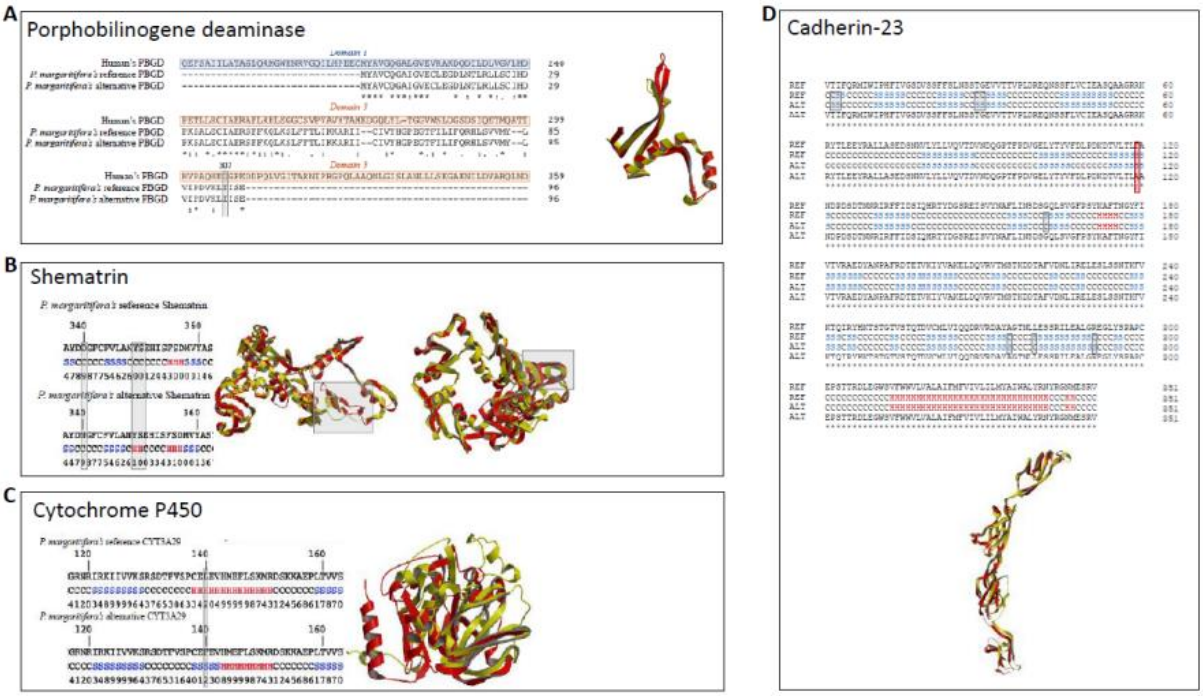


Figure 6

Protein modelling for the reference (genome sequence) and alternative pearl oyster sequence (carrying a SNP). A – Modelling for the porphobilinogen deaminase (PBGD): ClustalW alignment of the reference and alternate sequence of *P. margaritifera* (gray box highlights the SNP position) along with the superimposition of the Protein Data Base files for reference (red) and alternative (yellow) sequences. B - Modelling for the shematrin: ClustalW alignment of the reference and alternate sequence of *P. margaritifera* and predicted secondary structure (gray boxes highlight the SNP position (left) and the secondary structure modification (right)), along the superimposition of both reference (red) and alternative (yellow) shematrin Protein Databank file (gray boxes locate the differences on the superimposition). C - Modelling for the CYT3A29: ClustalW alignment of the reference and alternate sequence of *P. margaritifera*, predicted secondary structure (gray box highlights both the SNP position and the secondary structure modification), along with the superimposition of both reference (red) and alternative (yellow) CYT3A29 Protein Databank file. D - Modelling for the cadherin-23: ClustalW alignment of the reference and alternate sequence of *P. margaritifera* and their corresponding secondary structure (gray boxes highlight the secondary structure modification, and the red box shows the SNP position), along the superimposition of both sequences reference (red) and alternative (yellow).


A
Porphobilinogene deaminase
B
Shematrin
C
Cytochrome P450
D
Cadherin-23