



## Article

# Balanced Polymorphism at the *Pgm-1* Locus of the Pompeii Worm *Alvinella pompejana* and Its Variant Adaptability Is Only Governed by Two QE Mutations at Linked Sites

Alexis Bioy <sup>1</sup>, Anne-Sophie Le Port <sup>1</sup>, Emeline Sabourin <sup>2</sup>, Marie Verheye <sup>3</sup>, Patrice Piccino <sup>4</sup> , Baptiste Faure <sup>5</sup>, Stéphane Hourdez <sup>6</sup>, Jean Mary <sup>1</sup> and Didier Jollivet <sup>1,\*</sup> 

<sup>1</sup> Adaptation et Diversité en Milieu Marin, Equipe ABICE, Station Biologique de Roscoff, UMR 7144 Sorbonne Université-CNRS, 29688 Roscoff, France; alexis.bioy@hotmail.fr (A.B.); asleport@sb-roscoff.fr (A.-S.L.P.); jmary@sb-roscoff.fr (J.M.)

<sup>2</sup> Institut de Recherche Pour la Conservation des Zones Humides Méditerranéennes, Tour du Varlat, Le Sambuc, 13200 Arles, France; emeline.sabourin@gmail.com

<sup>3</sup> Royal Belgian Institute of Natural Sciences, Rue Vautier 29, 1000 Brussels, Belgium; mverheye@naturalsciences.be

<sup>4</sup> Lycée Lakanal, 3 Avenue du Président Franklin Roosevelt, 92330 Sceaux, France; patricepiccino@yahoo.fr

<sup>5</sup> Biotope—Agence Nord-Littoral, ZA de la Maie, Avenue de l'Europe, 62720 Rinxent, France; bfaure@biotope.fr

<sup>6</sup> Observatoire Oceanologique de Banyuls-sur-Mer, UMR 8222 CNRS-SU, 1 Avenue Pierre Fabre, 66650 Banyuls-sur-Mer, France; hourdez@obs-banyuls.fr

\* Correspondence: jollivet@sb-roscoff.fr; Tel.: +33-298292367; Fax: +33-298292324



**Citation:** Bioy, A.; Le Port, A.-S.; Sabourin, E.; Verheye, M.; Piccino, P.; Faure, B.; Hourdez, S.; Mary, J.; Jollivet, D. Balanced Polymorphism at the *Pgm-1* Locus of the Pompeii Worm *Alvinella pompejana* and Its Variant Adaptability Is Only Governed by Two QE Mutations at Linked Sites. *Genes* **2022**, *13*, 206. <https://doi.org/10.3390/genes13020206>

Academic Editor: Arcadi Navarro Cuartiellas

Received: 19 November 2021

Accepted: 17 January 2022

Published: 24 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Abstract:** The polychaete *Alvinella pompejana* lives exclusively on the walls of deep-sea hydrothermal chimneys along the East Pacific Rise (EPR), and displays specific adaptations to withstand the high temperatures and hypoxia associated with this highly variable habitat. Previous studies have revealed the existence of a balanced polymorphism on the enzyme phosphoglucosmutase associated with thermal variations, where allozymes 90 and 100 exhibit different optimal activities and thermostabilities. Exploration of the mutational landscape of phosphoglucosmutase 1 revealed the maintenance of four highly divergent allelic lineages encoding the three most frequent electromorphs over the geographic range of *A. pompejana*. This polymorphism is only governed by two linked amino acid replacements, located in exon 3 (E155Q and E190Q). A two-niche model of selection, including 'cold' and 'hot' conditions, represents the most likely scenario for the long-term persistence of these isoforms. Using directed mutagenesis and the expression of the three recombinant variants allowed us to test the additive effect of these two mutations on the biochemical properties of this enzyme. Our results are coherent with those previously obtained from native proteins, and reveal a thermodynamic trade-off between protein thermostability and catalysis, which is likely to have maintained these functional phenotypes prior to the geographic separation of populations across the Equator about 1.2 million years ago.

**Keywords:** phosphoglucosmutase; balancing selection; thermal stability; gene; adaptive mutations; Alvinellidae



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

A central goal in evolutionary biology is to understand the origin and maintenance of polymorphisms sculpted by natural selection and, more specifically, how the mean phenotype of a population evolves under heterogeneous and/or changing conditions [1]. As a consequence, many studies have investigated the maintenance of enzyme polymorphisms by selective processes for species exposed to environmental gradients such as temperature, salinity, or desiccation [2]. A few decades ago, a series of enzymes interacting in the glycolytic cycle, mostly associated with isomerase and mutase functions, such as phosphoglucose isomerase (PGI), mannose phosphate isomerase (MPI), and phosphoglucosmutase

(PGM), were shown to display isoforms that may be the subject of natural selection, leading to habitat-driven differentiation in populations according to temperature, wave action, or metallic pollution [2–10]. According to Eanes [11], such branch-point enzymes—which are positioned at the crossroad of metabolic pathways—are likely to be the target of natural selection, as they can orient pathway fluxes according to their protein variation. Among these, alleles encoding the enzyme phosphoglucomutase have been widely studied, with the aim of testing the hypothesis of differential and/or balancing selection. This was mainly achieved by looking at allele [3,4,12] and heterozygote [13] frequencies in populations, as well as assessing either the fitness of individuals carrying alleles suspected to be locally advantageous along latitudinal clines [14] or the kinetic properties of the enzyme isoforms themselves [6,12].

Due to their tremendous thermal variability, caused by the chaotic mixing of cold sea water and hot fluids, hydrothermal vents represent an ideal model for testing the effect of frequent, and unpredictable, spatial and temporal changes of habitats on adaptive enzyme polymorphisms. First, both the fragmentation and instability of the vent discharge likely promote highly dynamic meta-populations with recurrent local extinctions and associated bottlenecks [15–17]. Long-term oscillations of heat convection beneath the ridge lead to the displacement of the hydrothermal activity along the rift, generating the emergence of new vent sites more or less close to older ones that became extinct, allowing for their rapid recolonization [16,18]. Such dynamics should have severe implications in reducing the genetic diversity of vent species. Second, variations in temperature, sulphide concentrations, and oxygen concentrations over short periods of time (often ranging from minutes to hours) [19,20] are likely to affect the respiratory, nutritional, and reproductive physiologies of animals living in such places [21]. Animals must be able to feed under high temperatures to fuel their symbionts and respire under cold conditions to obtain oxygen. These temporal fluctuations of the vent conditions at the individual scale represents a selective constraint that should promote the maintenance of both thermostable and cold-functioning enzyme alleles through mean overdominance until the exploration of the mutational landscape of a given enzyme leads to the emergence of a highly plastic allelic isoform, and thus enzyme monomorphism. However, the hydrothermal environment is also highly fragmented and heterogeneous, according to the mineral composition of the oceanic crust through which the super-heated fluid moves prior to be expelled above the seafloor [22]. As a consequence, vent fields often display a mosaic of edifices of different ages [23,24], whose mean age and size distribution is dictated by the frequency of tectonic and volcanic events, as well as the dynamics of the heat convection beneath oceanic ridges [25–27]. Vent edifices usually cool down with age until they turn off. Depending on the age of the edifice, populations of vent species are thus spatially subjected to a wide variety of vent conditions, which represents an ecological basis for diversifying selection. As most vent species are quite thermotolerant, such a mosaic of habitats is likely to favour thermostable alleles in newly formed edifices and cold functioning ones in older habitats, the fate of a given allele depending on the relative proportions of these habitats.

The polychaete *Alvinella pompejana*, which lives on the hottest part of the hydrothermal-vent environment [28,29] can withstand temperatures up to 50 °C [30]. This tube-dwelling worm lives on the walls of hydrothermal-vent chimneys from a latitude of 23° N on the East Pacific Rise (EPR) to 38° S on the Pacific Antarctic Ridge (PAR) [31]. It has developed peculiar physiological adaptations in order to colonize this hostile habitat [32,33]. Earlier genetic studies have shown that *A. pompejana* exhibits quite an unusually high level of genetic diversity [31,34,35] with a non-negligible number of bi-allelic enzyme loci with equally frequent alleles, some of which display different thermal stabilities [36]. Among them, the enzyme phosphoglucomutase (PGM-1) possesses four distinct isoforms. Allozymes 90 and 100 have frequencies of approximately 35% and 60%, respectively, in populations of the northern EPR, while the two other isoforms (112 and 78) are rather rare, accounting for the remaining 5%. Although the frequency of allozyme 90 remains constant over the species range, Plouviez et al. [35] have shown that allozymes 78 and 100

display an abrupt clinal distribution across the Equator, with allozyme 78 becoming the most frequent allele in the southern EPR. Bi-allelism was, thus, preserved all along the EPR despite population isolation, recurrent extinction/recolonizations, and a long history of divergence across the Equatorial barrier.

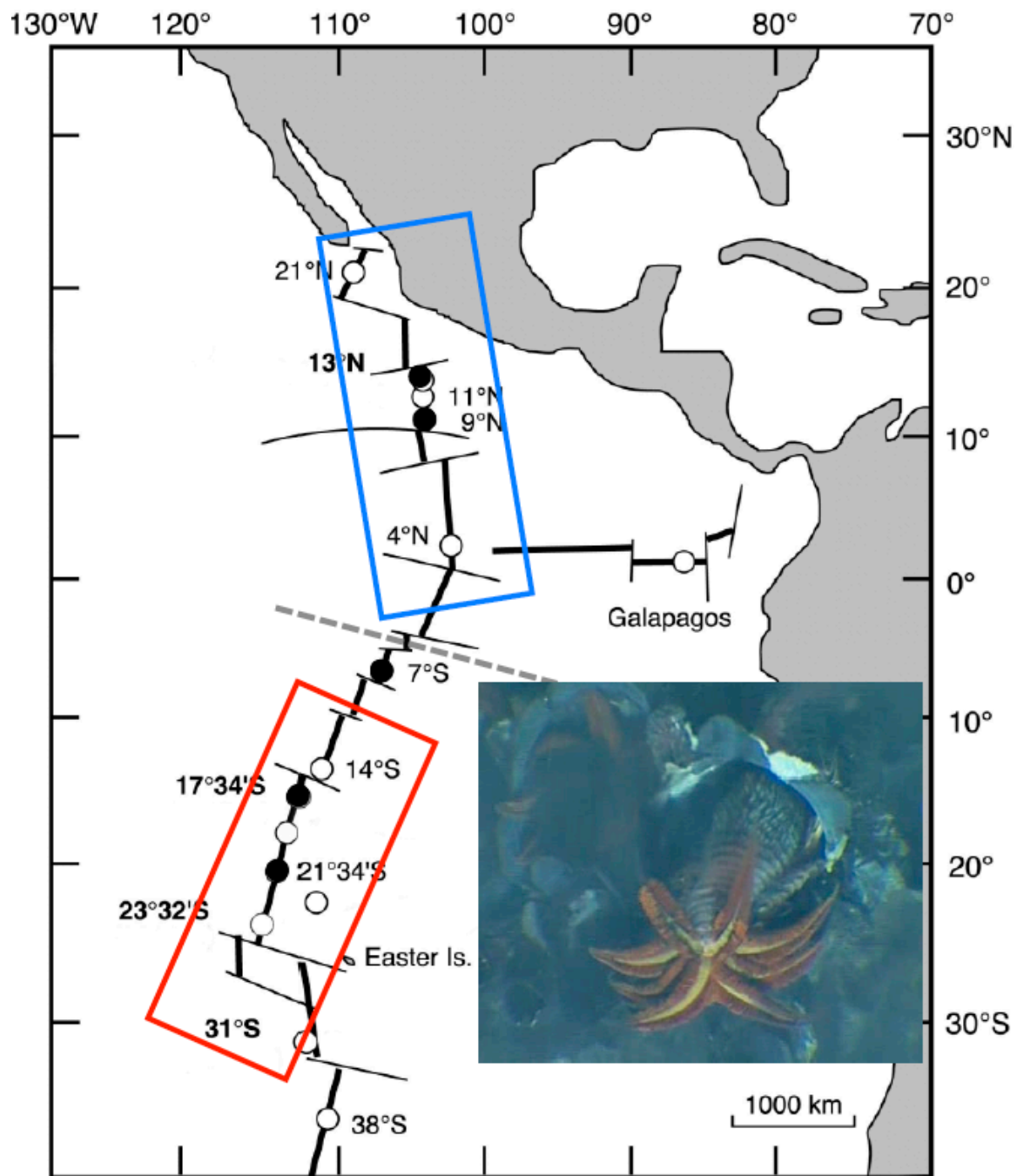
In addition, significant genetic differentiation has been observed between *A. pompejana* populations living in contrasting microhabitats, especially when comparing newly formed ‘still hot’ chimneys (i.e., ‘hot’ niche) to older and colder edifices (i.e., ‘cold’ niche). The frequency of allele 90 is, indeed, positively correlated with mean temperature at the opening of *Alvinella* tubes and increases in the ‘hot’ compared to the ‘cold’ habitat, suggesting that this locus is under diversifying selection [12]. In vitro experiments on enzyme stability and optima have strengthened this view. They have shown that allele 90 is more thermostable and more active at higher temperatures than allele 100 and, thus, is probably favoured in the ‘hot’ habitat.

Although whole-length *Pgm-1* sequences have now been obtained for a large panel of metazoan species, very few studies have been conducted at the population level, most of which involved bacterial strains. While this enzyme has been extensively studied from the 1970s to 1990s for adaptive purposes, only few studies have examined the relationship between nonsynonymous changes at the gene level and the subsequent enzymatic performance of alternate isoforms (see, e.g., [14,37] for the correspondence between allelism, enzyme thermal resistance, and glycogen storage in *Drosophila*). In this paper, we report a possible case of long-term balancing selection at an enzyme locus where alleles can be maintained by varying selection between two niches, possibly helped by fluctuations in the relative proportions of the two niches over space and time. Most of the documented cases for the long-term persistence of alleles by balancing selection and trans-species polymorphism come from studies dealing with negative frequency-dependent selection at immune and sex-determination genes [38,39]. This raises questions about how a chaotic and highly fluctuating two-niche system can promote balancing selection at key branch-point enzymes. The aim of this work was, therefore, to identify the mutations at the basis of the enzyme polymorphism of PGM-1 in the Pompeii worm and to evaluate the effects of these mutations on the thermostability and catalytic efficiency of the enzyme to test whether a trade-off between these two processes is likely to explain the maintenance of the different isoforms. In parallel, we explored the mutational landscape of the gene to search for traces of balancing selection in the vicinity of the amino acid substitutions that led to these isoforms and assess their long-term duration. Finally, we also examined the evolutionary mechanisms: (1) fitness cost to the colonization of newly ‘hot’ habitats; (2) overdominance associated with the fluctuations of the vent discharge, and (3) the dynamics of hot vs cold habitats (i.e., the two-niches model hypothesis), by which these isoforms could have been maintained in natural populations by analyzing the fecundity of females carrying the most thermostable isoform, and by using MSMS simulations implementing selection across a barrier to gene flow.

## 2. Materials and Methods

### 2.1. Animal Sampling

Specimens of *A. pompejana* were collected with the remotely operated vehicle (ROV) Victor 6000 and the deep-sea manned submersible Nautilie during the cruises Phare 2002, Biospedo 2004, and Mescal 2010 on board the research vessel L’Atalante. Animals were sampled from targeted sites located on the northern EPR (NEPR) and southern EPR (SEPR; see Figure 1) over chimneys of different ages ranging from newly formed ‘hot’ diffusers to large black ‘smokers’, at which the thermal and chemical conditions were highly contrasting [24].



**Figure 1.** Species range of the Pompeii worm *Alvinella pompejana* along the East Pacific Rise (EPR). Dashed line indicates the presence of the Equatorial barrier to gene flow depicted by Plouviez et al. (2009, 2010) [35,40]. Blue and red boxes correspond to the northern and southern metapopulations of the worm. Dots represent localities where the worms are present and black ones, the localities from which the worms have been sampled to perform the present study. Inset includes a close 10 cm<sup>2</sup> view of *A. pompejana* moving outside its tube on the wall of a hydrothermal vent chimney (IFREMER copyright). The photograph shows the anterior part of the worm with its branchial crown and its epibiotic filamentous bacteria located dorsally.

In order to test an earlier hypothesis [12], postulating that individuals carrying genotypes favoured during the colonization of newly formed still ‘hot’ chimneys may be counter-selected by a lower reproductive fitness under cooler conditions (i.e., trade-off between colonization ability and reproduction), we examined the relationship between

*Pgm-1* genotypes and the fecundity of females. To perform this test, the size of animals was estimated from the width at the S4 setigerous segment, and sexes were determined based on the presence of either a genital pore in females or a pair of sexual tentacles in males [41]. Mature females collected from both sides of the Equator were dissected to estimate their fecundity per size unit and genotyped at the *Pgm-1* locus. For each female, the coelomic fluid containing oocytes was carefully removed and re-suspended in 50 mL of a solution of borate-buffered 3% formalin in seawater. Oocytes were counted following the method previously described by Faure et al. [42], and one-way ANOVA was performed on size-corrected female fecundities, according to the genotype, using the Jamovi software (<https://www.jamovi.org> (12 March 2020)).

## 2.2. Identification and Characterization of the AP-*Pgm-1* Gene

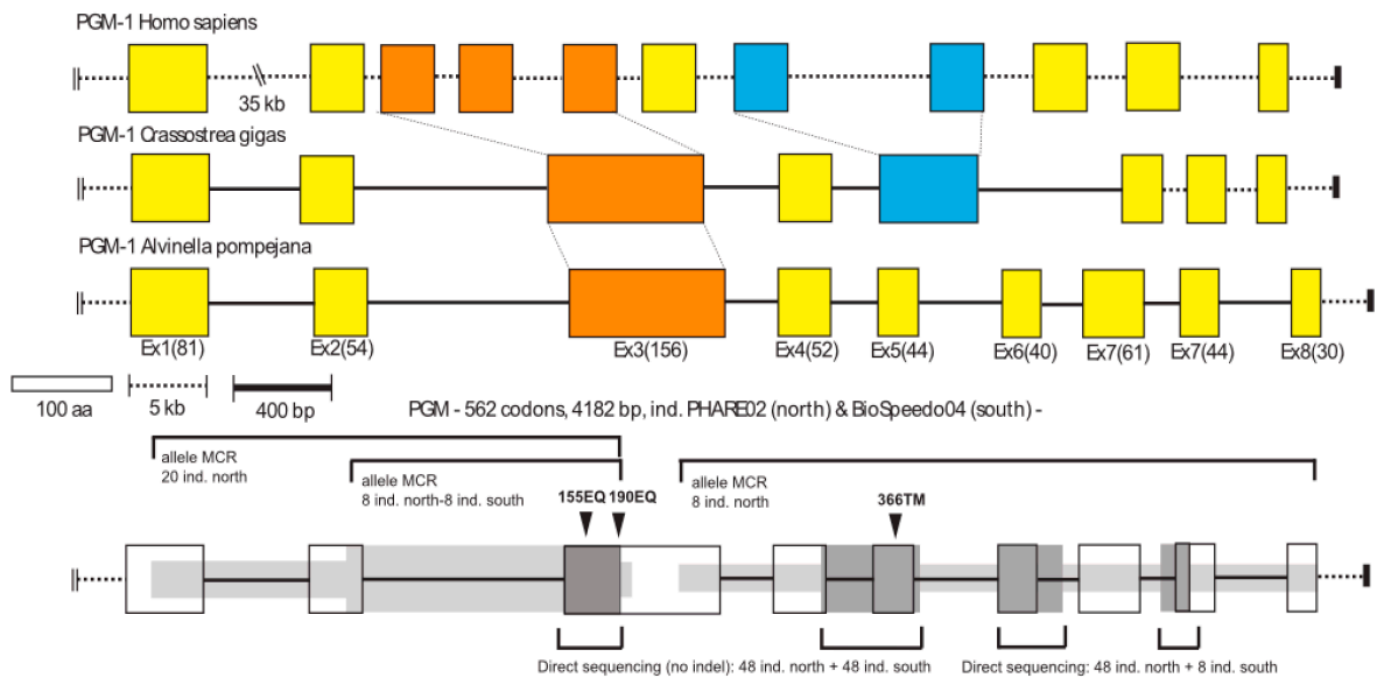
### 2.2.1. Sequencing of AP-*Pgm-1* cDNA Using Homozygous Individuals

Based on allozyme genotypes, eight homozygous individuals carrying alleles 78, 90, and 100 were selected for RNA extractions. Total RNAs were extracted with Tri-Reagent (Sigma-Aldrich, St Quentin-Fallavier, France), following the manufacturer's instructions, and a classical chloroform extraction protocol. Both the quantity and quality of RNA were assessed using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Villebon-sur-Yvette, France). Five µg of total RNAs were reverse transcribed with M-MLV reverse transcriptase (Promega, Charbonnières-les-bains, France), anchor-oligo(dT) primer (Table S1), and random hexamers (Promega). Reverse anchor and forward-nested degenerate *Pgm* primers derived from the oyster *Crassostrea gigas* and human *Pgm-1* sequences were then used to perform a preliminary amplification of cDNA fragments (see Table S1). PCR-products containing *Pgm-1* cDNA candidates were then cloned using a TOPO TA Cloning kit (Invitrogen, Villebon-sur-Yvette, France) and sequenced on an ABI 3130 sequencer using the BigDye v.3.1 (PerkinElmer, Villebon-sur-Yvette, France) terminator chemistry, following the manufacturer's protocol. Sequences of clones containing appropriate *Pgm-1* cDNA fragments were then aligned to reconstruct a series of nearly complete AP-*Pgm-1* cDNA (i.e., only lacking a small part of the 5'-end of the coding sequence).

### 2.2.2. Sequencing the AP-*Pgm-1* Gene with a Series of Specific Exonic Primers

Using gDNA, specific reverse primers (Table S1) were also used to amplify the 5' portion of the gene by directional genome walking using PCR [43]. A series of specific primers were designed based on previously obtained cDNA sequences (see Table S1) in order to amplify both exon- and intron-containing portions of the gene using the same gDNA from the eight homozygous individuals. Fragments of the gene were obtained using pairs of the least distant forward and reverse primers containing a 6-bp individual identifier (barcode). PCR amplifications were performed in a 25 µL PCR reaction volume comprising 1× buffer (supplied by manufacturer), 2 mM MgCl<sub>2</sub>, 0.25 mM of each dNTP, 0.4 µM of each primer, and 0.5 U of Taq polymerase (Thermoprime plus). The PCR profile included a first denaturation step at 94 °C for 4 min, followed by 30 cycles at 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 2 min, then a final extension at 72 °C for 10 min. All barcoded PCR products were cloned following the Molecular Cloning Recapture (MCR) method developed by Bierne et al. [44] and sequenced on an ABI 3130 sequencer with the protocol used previously. Alignments of the sequenced fragments allowed us to reconstruct a complete sequence of the AP-*Pgm-1* gene (Accession Number MN218831), its associated cDNA sequence, and three native consensus cDNA for the three isoforms (Accession Numbers MN218832–MN218839). The analysis of this initial cDNA alignment provided initial information on polymorphic sites between the three distinct alleles along the gene (see Figure 2).





**Figure 2.** Gene organization of the *Pgm-1* gene for human (*Homo sapiens*), the oyster (*Crassostrea gigas*) and the worm *A. pompejana*. Scale bars indicate the size of exons (white box in terms of amino-acid residues) and introns (dashed and black bars: dashed scaled bar was more appropriate for the human gene which has introns of greater size). Values in parentheses correspond to the exact number of amino acid residues per exon in the worm *A. pompejana*. The dark and light grey zones associated with the *AP-Pgm1* gene (bottom of the figure) highlight the different regions of the gene which has been sequenced for the study with the Mark, Cloning, Recapture (MCR) method (light grey) or direct sequencing (dark grey) together with the number of individuals used. Black arrows indicate the position of the main amino-acid substitutions onto the gene.

### 2.2.3. Correspondence between Allozymes and Non-Synonymous Mutations of *AP-Pgm-1*

To examine the correspondence between the only two diagnostic polymorphic non-synonymous EQ mutations found at exon 3 and allozymes 78, 90, and 100, a total of 220 individuals were genotyped on the 350 bp fragment of the *Pgm-1* exon 3 containing these sites. PGM-1 allozymes were first screened for each individual by electrophoresis on 12% starch-gel at 4 °C (100 V, 80 mA, 4 h) with the Tris-citrate pH 8.0 buffer system, following the procedure described by Piccino [12]. The 350 bp exon 3 fragment was then amplified by PCR for the same individuals, following gDNA extraction using the CTAB/PVP procedure described by Plouviez [40]. PCR amplifications were conducted using a specific primer pair (see Table S1) with a first denaturation step at 94 °C for 4 min, followed by 40 cycles at 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 20 s, then a final extension at 72 °C for 2 min. PCR products were first double-digested on 33 individuals using the enzymes *Fai I* (targeting the first substitution site) and *Bsg I* (targeting the second site) as an initial test, and then sequenced without cloning using an ABI 3130 automatic sequencer with the BigDye v.3.1 (PerkinElmer) terminator chemistry after an ExoSAP-IT purification.

Forward and reverse sequences were proof-read in CodonCode Aligner to check for the occurrence of single (homozygotes) or double (heterozygotes) peaks at the two polymorphic sites. The allele alignment was deposited in Genbank (accession numbers MN218918–MN219291). Linkage disequilibrium between genotypes EE, EQ, and QE and allozymes 78, 90, and 100 was tested using Linkdis function [45] of the Genetix v.4.05 software [46]. The double mutation scoring among individuals allowed us to then estimate heterozygote excesses or deficiencies within populations. Departures to HWE were tested with 1000 permutations of alleles between genotypes using the same software. The exon 3 allele alignment was also used to reconstruct an allelic network, using Network 4.5.1.0 [47],

in order to examine gene flow across the equatorial barrier between populations at this locus.

#### 2.2.4. Examining the Synonymous and Non-Synonymous Changes along *AP-Pgm-1*

Nucleotidic diversities were punctually assessed along the gene by combining direct sequencing and the MCR method, considering individuals from each side of the EPR (see Figure 2). These regions included exon 1, exons 4 to 5, the end of exon 7, and the beginning of exon 9 (Accession Numbers MN218840–MN218917 for exon1, Accession Numbers MN219292–MN219356 for exons 4 and 5, see Figure 2). In addition, a fragment containing the whole intron 2 and the beginning of exon 3, where the two diagnostic EQ mutations are located (1110 bp), was also sequenced using the MCR method [44]. This allowed us to test whether mutation hotspots occur around these two EQ sites and to estimate allele divergence (Accession Numbers MN219357–MN219404). In the MCR sequence sets, the number of retrieved alleles greatly varied between the different parts of the gene according to the sequencing efficiency and/or cloning success. Artefactual singletons due to the MCR method were removed by comparing singleton rates between the MCR and direct sequencing data sets on the same fragments.

Haplotype diversity ( $H_d$ ), nucleotide diversity ( $\pi$ ) and its synonymous and non-synonymous components ( $\pi_S$  and  $\pi_N$ ), and Watterson's theta ( $\theta_W$ ) were then examined together with deviations to neutral evolution (Tajima's D and Fu & Li's F statistics) for both the northern and southern EPR individuals along the gene, using the DNAsp 4.10.3 software [48] with a sliding window (size = 100 and step = 25). These basic genetic parameters were then compared with the critical values associated with sample size and neutral coalescent simulations, as implemented in the same software. Linkage disequilibrium between segregating sites and recombination among alleles were estimated by calculating the  $Z_{nS}$  statistics [49] together with the minimum number of recombination events ( $R_m$  [50]). The number of significant associations between linked sites was evaluated following a Fisher's exact test and a Bonferroni correction, implemented in DNAsp 4.10.3. The occurrence of recombinants was also checked using automated RDP and bootscan packages of RDP v.3.44 [51], and recombination hotspots were searched by examining the population recombination rate parameter ( $4N_e.r$ ) along the gene (-recomb and -hotspot outputs) for both the northern and southern populations using Phase 2.1.1 software [52]. Genetic differentiation and allele divergence between the southern and northern parts of the EPR were estimated by calculating  $F_{st}$  and  $D_{xy}$  in DNAsp 4.10.3. Genetic differentiation was tested using 1000 permutations of the sequence data sets using the randomization test developed by Hudson [53]. Finally, the intron 2-exon 3 alignment (1110 bp) was used to reconstruct a coalescent tree of *AP-Pgm-1* alleles to more specifically evaluate both the intra-locus recombination and allele divergence near the two EQ non-synonymous polymorphic sites. The evolutionary history of alleles was inferred using the Minimum Evolution method implemented in MEGA7 [54] using the NJ algorithm for the initial tree, pairwise deletion of ambiguous sites, and the close-neighbour-interchange (CNI) algorithm. Evolutionary distances were computed using the Maximum Composite Likelihood method.

#### 2.3. Coalescent Simulations Using Models of Selection

Coalescent simulations under a divergence model with asymmetrical migration rates between two populations ( $pop_{1 \rightarrow 2}: 2N_e.m = 1$  and  $pop_{2 \rightarrow 1}: 2N_e.m = 0.1$ ,  $N = 1000$  simulations with  $N_e = 50,000$ ) were performed using MSMS v3.2 software [55]. Two different hypotheses of balancing selection were evaluated: (1) over-dominance, with genotype selection coefficients  $S_{aA} = 500$  and  $S_{AA} = 1$  where S represents  $N_e.s$ ; and (2) habitat-dependent selection, with four populations and two habitats, where  $S_{AA} = 1000$ ,  $S_{aA} = 500$ , and  $S_{aa} = 0$  in the first habitat and  $S_{AA} = 0$ ,  $S_{aA} = 500$ , and  $S_{aa} = 1000$  in the second habitat. Each set of simulations, including the null hypothesis of asymmetrical migration without selection, were run with two recombination rates ( $N_e.r = 1$  or 100). Population parameters, including

gene diversities ( $\theta_W$ ,  $\pi$ ), and Tajima's D within each deme and  $F_{st}$  between demes were estimated using the pylibseq 0.2.3 libraries [56] and a home-made python script.

#### 2.4. Functional and Structural Analysis of PGM-1 Recombinant Isoforms

##### 2.4.1. Plasmid Construction for Enzyme Expression

Full-length *AP-Pgm* cDNA was obtained from two homozygous individuals: 100/100 (EE) and 90/90 (EQ). RT-PCR was conducted using a ClonTech SMARTer Race cDNA amplification kit, following the manufacturer's instructions and using AP-PGMex11 reverse primer (see Table S1). These cDNAs were then used as a target to specifically amplify the complete coding sequence, with primers containing cutting sites to be inserted in either Pet20b or PetDuet expression vectors (Table S1). Amplified coding sequences were double digested with either the enzymes *Bam*HI/*Not*I or *Ase*I/*Xho*I, in a 25  $\mu$ L volume containing a restriction buffer, the enzymes, and 1% BSA. The restriction products were then ligated in the appropriate expression vector after purification with a Nucleospin Gel extraction Clean up column (Macherey Nagel, Hoerd, France) and cloned into BL21DE3 *Escherichia coli* cells amenable for IPTG induction and protein expression.

##### 2.4.2. Directed Mutagenesis

Using full-length cDNA with the double mutation EE as a template, mutants  $_{155}$ EQ and  $_{190}$ EQ were produced by directed mutagenesis following the PCR protocol of Reikofski and Tao [57]. First, amplifications were conducted in a 50  $\mu$ L reaction volume containing  $1 \times$  Pfu buffer containing  $MgCl_2$ , 0.25 mM of each dNTP, 0.5  $\mu$ M of each primer (petDuet and mutated primer), and 0.5 U of the proof-reading *Pfu* polymerase (Promega) with 30 cycles of 94  $^{\circ}$ C for 30 s, 60  $^{\circ}$ C for 30 s, and 72  $^{\circ}$ C for 3 min. Then, the two regions of the mutated cDNA were joined following PCR amplification without primers, mixing the two previous PCR-products (1:1) under the same conditions and with a final elongation step of 10 min. cDNAs containing the mutated sites  $_{155}E \rightarrow Q$  and  $_{190}E \rightarrow Q$  and the native  $_{155}E_{190}E$  cDNA were then sequenced using an ABI 3130 sequencer with the BigDye v.3.1 (Perkin Elmer) terminator chemistry to verify the sequences before protein expression.

##### 2.4.3. Protein Expression and Purification

*E. coli* (BL21DE3) with the recombinant pETduet plasmid containing either native or mutated *Pgm-1* cDNA sequences were grown in LB medium supplied with 100  $\mu$ g/mL ampicillin at 37  $^{\circ}$ C until they reached an absorbance of 0.6 at 600 nm. Protein expression was induced by adding 1 mM IPTG to the medium and kept at 37  $^{\circ}$ C under shaking for 4 h. Cells were then harvested by centrifugation (4  $^{\circ}$ C/15,000  $\times$  g/5 min), and the pellets were resuspended in a binding buffer (20 mM Tris-HCl, pH 6.5; 500 mM NaCl; 5 mM imidazole), disrupted by French Press at 1.6 kbar. After removing cell debris by centrifugation (15,000  $\times$  g/4  $^{\circ}$ C/60 min), the supernatants (1  $\mu$ g/mL of lysate) were treated with DNase I (Eurogentec, Angers, France) for 1 h on ice. A first purification step was performed using immobilized metal affinity chromatography with a His-bind resin column (His-Bond kit, Novagen<sup>®</sup>, Molsheim, France) to recover PGM-1 variants. Protein binding with 5 mM and 60 mM imidazole and final elution of allozymes with 1 M imidazole were performed, following a classical chromatography protocol (pH 6.5). The eluted fractions were concentrated using 30 kDa molecular cut-off Amicon-Ultra (Millipore<sup>™</sup>). A second purification step was performed by size-exclusion chromatography (SEC) with a Superdex 75 column (1  $\times$  30; GE Healthcare, Velizy, France) at a flow rate of 0.5 mL/min, monitored at 280 nm using 25 mM  $Na_2HPO_4/NaH_2PO_4$ , pH 6.5. The purity of proteins was checked by SDS-PAGE stained with Coomassie brilliant blue, before being kept at 4  $^{\circ}$ C in elution buffer supplemented with dithiothreitol (DTT, 10 mM) until use for enzyme assays. The protein concentrations were measured by absorption at 280 nm, with the theoretical coefficient of 48,820  $M^{-1}.cm^{-1}$ , as calculated using the ExPASy-ProtParam tool (<http://web.expasy.org/protparam/> (20 June 2019)).



#### 2.4.4. Enzyme Activity Assay

PGM activities were assayed by coupling the formation of  $\alpha$ -D-glucose 6-phosphate (G6P) from  $\alpha$ -D-glucose 1-phosphate (G1P) to NADPH formation using glucose 6-phosphate dehydrogenase (G6PD) as a relay enzyme. The reaction mixture contained 50 mM Tris-HCl, pH 7.4; 0.5 M  $MgCl_2$ ; 1.2 mM NADP; and 0.1  $\mu$ M G6PD. The recombinant PGMs were used at the following concentrations: [PGM<sub>78</sub>] = 0.9  $\mu$ M, [PGM<sub>90</sub>] = 4  $\mu$ M, and [PGM<sub>100</sub>] = 0.6  $\mu$ M. The concentration of the substrate (G1P) was varied from 0.2 to 60 mM, to determine the kinetic constants  $K_m$  and  $V_{max}$  using a Lineweaver–Burk plot.

#### 2.4.5. Thermal Inactivation

The purified PGM activities were measured at 37 °C and at 340 nm using an UVmc<sup>2</sup> spectrophotometer (Safas, Monaco) after a 30-min incubation at challenge temperatures ranging from 5 to 60 °C. Activities were then normalized as the percentage of residual activity when compared to the same sample kept on ice. A theoretical curve with the following equation was fitted to each experimental data set using a nonlinear curve fit algorithm (Kaleidagraph 4.5.0, Synergy Software, Reading, PA, USA) [58]:

$$y = \frac{(y_N + m_N \cdot T) + (y_D + m_D \cdot T) \cdot \exp\left(\frac{m(T - T_m)}{RT}\right)}{1 + \exp\left(\frac{m(T - T_m)}{RT}\right)}, \quad (1)$$

where  $y$  is the residual activity;  $y_N$ ,  $m_N$ ,  $y_D$ , and  $m_D$  are parameters characterizing the activity of the native enzyme (N) and its denatured form (D), respectively;  $m$  characterizes the transition between the native and denatured forms;  $R$  is the universal gas constant;  $T$  is the absolute temperature, and  $T_m$  the absolute temperature of half-denaturation (i.e., the temperature at which the activity of the enzyme is reduced by half).

#### 2.5. Guanidinium Chloride-Induced Unfolding of PGM Isoforms

Unfolding of PGM isoforms was induced by guanidinium chloride ( $GdmCl$ ) in a 25 mM sodium phosphate (pH 6.5) and 200 mM NaCl buffer. Proteins (12  $\mu$ M) were incubated with increasing concentrations of  $GdmCl$  from 0 to 5 M, for 30 min at 20 °C, and their intrinsic fluorescence emission was determined at 324 nm under excitation at 290 nm using a Safas Xenius spectrofluorimeter (Safas, Monaco). The  $GdmCl$  concentration was determined by refractive index measurements [59]. Biphasic states of protein denaturation, with an intermediate state (I) between native (N) and unfolded (U) states according to the equilibrium  $N \leftrightarrow I \leftrightarrow U$ , were treated as follows. We assumed that each transition ( $N \leftrightarrow I$  and  $I \leftrightarrow U$ ) followed a two-state model of denaturation. The denatured protein fraction for each transition— $f(I)$  for transition ( $N \leftrightarrow I$ ) and  $f(II)$  for transition ( $I \leftrightarrow U$ )—was determined by resolving the two following equations:

$$f(I) = (y_N - y)/(y_N - y_I),$$

$$f(II) = (y_I - y)/(y_I - y_U),$$

where  $y_N$ ,  $y_I$ , and  $y_U$  are the measured fluorescence intensities of the native, intermediate, and unfolded states, respectively, and  $y$  is the fluorescence intensity observed at a given  $GdmCl$  concentration. The unfolded fraction  $f(I$  or  $II)$  data were plotted against  $GdmCl$  concentrations and theoretical curves, defined by the following equation, were fitted to the experimental data set using a nonlinear curve fitting algorithm (Kaleidagraph 4.5.0, Synergy Software) [60]:

$$f(I \text{ or } II) = \frac{\exp\left(-m \frac{(C_m - [GdmCl])}{RT}\right)}{1 + \exp\left(-m \frac{(C_m - [GdmCl])}{RT}\right)}, \quad (2)$$

where  $T$  is the absolute temperature,  $R$  is the universal gas constant,  $C_m$  is the concentration of *GdmCl* at the mid-point of the transition, and  $m$  the dependence of the Gibbs free energy of the unfolding reaction ( $\Delta G$ ) on the denaturation concentration of *GdmCl*. Knowing  $C_m$  and  $m$ , the standard Gibbs free energy of the unfolding reaction in the absence of denaturant,  $\Delta G^0_{H_2O}$ , can be calculated according to the relation [58]:

$$\Delta G^0_{H_2O} = m C_m \quad (3)$$

### 2.6. 3D-PGM Structure Modelling

PGM 78, 90, and 100 3D protein conformations were modelled using Modeller 9v13 software [61], using the structure of the crystallized rabbit phosphoglucomutase with its substrate,  $\alpha$ -D-glucose 1-phosphate, as a template (pdb file 1C47). This protein comprises 561 amino acids with a resolution of 2.70 Å, sharing 65% sequence identity with that of *A. pompejana*. A total of 100 models were generated for each PGM isoform and their quality was assessed using the Modeller Objective Function parameter. Finally, structural optimization was obtained using the repair function of the FoldX software [62].

## 3. Results

### 3.1. Sequencing AP-Pgm-1 cDNA from Homozygous Genotypes

Full-length *Pgm-1* cDNA sequences were obtained from three 100/100 genotypes, three 90/90 genotypes, and only two 78/78 genotypes. This led to a complete cDNA sequence of 562 codons without indels between alleles encoding the three distinct allozymes (Figure S1). The consensus protein sequence fell into the phosphoglucomutase 1 family of proteins, with a blastp e-value of 0.0 (65–72% of identity over 99% of 562 residues with the sequence from the oyster *C. gigas* and a selection of vertebrate species). Of the 16 cloned sequences, only two nonsynonymous mutations on exon 3 allowed us to discriminate the three main allozyme genotypes (i.e., 100/100, 90/90, and 78/78). These polymorphic mutations corresponded to the replacement of a glutamic acid (E) by a glutamine (Q) at positions 155 and 190. Another replacement of a valine (V) by a leucine (L) at position 40 was also found in exon 1 at intermediate frequency, but this amino acid polymorphism was not linked to a given electromorph. The replacement of a phenylalanine (F) by a leucine (L) was also found at position 502 in cDNA encoding allozyme 90.

### 3.2. Assignment of the Two EQ Amino Acid Replacements to Allozymes in Natural Populations

To address the relationship between the two QE substitutions depicted by the cDNA sequences and allozymes, direct sequencing (and/or RFLP) was performed on a portion of exon 3 (94 codons) containing the double diagnostic mutations EQ in 220 individuals from both sides of the EPR, previously genotyped at the PGM-1 enzyme. The linkage disequilibrium between the two EQ mutations (at codon positions 155 and 190) and allozymes was highly significant (Table 1), with correlation coefficients ( $R_{ij}$ ) greater than 70% ( $p$ -values < 0.0001). This provided very reliable correlations, in which the combinations QE, EQ, and EE corresponded to the isoforms 78, 90, and 100, respectively. The most negatively charged allozyme 112, which is rare and always found at the heterozygous state in the northern populations, was also assigned to genotype EE, suggesting that an additional replacement occurs elsewhere in this protein. From this genotyping, groups of individuals from either the northern or the southern EPR did not depart significantly from the Hardy-Weinberg proportions. However, observed and expected heterozygosities were both greater in the northern population ( $H_{o-North}$ : 0.40 vs.  $H_{o-South}$ : 0.29). Interestingly, the allele QQ was not found in any of the populations. The frequencies of EE, EQ, and QE alleles at the sampled localities are summarized in Table S2. A more thorough analysis of the north/south genetic differentiation was conducted on the 374 allelic sequences obtained by direct sequencing (see alignment in Supplementary Data). The resulting haplotype network (Figure S2) shows a quasi-complete isolation of the northern and southern populations, with a  $F_{st}$  value of 0.510 (see Table 2). Based on the 282 bp alignment, PGM90 (EQ) found

in the southern population derives directly from the northern PGM90 (EQ) by one fixed mutation, while the southern PGM78 (QE) differs by two mutations from the northern PGM100 (EE). The haplotype network also indicated that at least three alleles sampled in the southern populations originated from the northern populations, suggesting that the barrier to gene flow is not complete.

**Table 1.** Linkage disequilibrium between the combination of the two diagnostic mutations EQ and PGM-1 allozymes.

Mutation	D <sub>ij</sub>	R <sub>ij</sub>	Chi <sup>2</sup>	p-Value
EE-100	0.274	0.908	87.2	0.0001 ***
EQ-90	0.115	0.725	55.7	0.0001 ***
QE-78	0.357	0.907	87.2	0.0001 ***
EE-112	0.013	0.230	5.6	0.0178 *

\*:  $p = 0.05$ ; \*\*\*:  $p = 0.001$ .

**Table 2.** Gene diversities, population parameters, and neutrality tests along the *Pgm-1* gene for *A. pompejana* populations from the southern and northern EPR.

Statistics	E1 North	E1 South	E2-I2 North	E2-I2 South	E3 North	E3 South	E4-E5 North	E4-E5 South	E7-E9 North	E7-E9 South
Fragment length (bp)	273	273	1110	1110	278	278	803	803	576	576
N	38	40	36	12	156	218	20	45	62	12
H <sub>d</sub>	0.77 ± 0.06	0.39 ± 0.01	0.95 ± 0.03	0.85 ± 0.02	0.72 ± 0.03	0.76 ± 0.04	0.68 ± 0.10	0.88 ± 0.04	0.91 ± 0.03	0.98 ± 0.04
Overall $\pi$	0.0059 ± 0.0006	0.0017 ± 0.0005	0.0056 ± 0.0003	0.0087 ± 0.0005	0.0045 ± 0.0003	0.0070 ± 0.0006	0.0023 ± 0.0005	0.0045 ± 0.0005	0.0044 ± 0.0005	0.0068 ± 0.0010
$\pi_s$	0.0016 (62.6)	0.0008 (62.6)	0.0126 (58.7)	0.0404 (58.7)	0.0034 (48.7)	0.0167 (48.7)	0.0094 (88.1)	0.0164 (88.1)	0.0000 (27.8)	0.0000 (27.8)
$\pi_n$	0.0029 (210.4)	0.0020 (210.4)	0.0027 (205.3)	0.0056 (205.3)	0.0030 (170.3)	0.0028 (170.3)	0.0003 (307.9)	0.0012 (307.9)	0.0018 (110.2)	0.0028 (110.2)
S	8	5	26	26	16	18	5	10	24	13
$\theta_w$ (S)	0.0070 ± 0.0031	0.0043 ± 0.0022	0.0057 ± 0.0033	0.0078 ± 0.0042	0.0102 ± 0.0026	0.0102 ± 0.0026	0.0036 ± 0.0022	0.0058 ± 0.0024	0.0094 ± 0.0030	0.0076 ± 0.0035
Z <sub>n</sub> S	0.054 (1/15/1 <sup>B</sup> )	0.008 (0/1/0 <sup>B</sup> )	0.110 (43/325/19 <sup>B</sup> )	0.466 (46/153/0 <sup>B</sup> )	0.0093 (2/36/1 <sup>B</sup> )	0.0540 (11/66/7 <sup>B</sup> )	0.0226 (0/3/0 <sup>B</sup> )	0.0324 (3/21/0 <sup>B</sup> )	0.0261 (4/120/1 <sup>B</sup> )	0.1641 (5/36/0 <sup>B</sup> )
R <sub>m</sub>	1 (RDP n.d.)	0 (RDP n.d.)	4 (RDP = 1)	0 (RDP n.d.)	3 (RDP n.d.)	6 (RDP n.d.)	0 (RDP n.d.)	3 (RDP n.d.)	5 (RDP n.d.)	2 (RDP n.d.)
F <sub>st</sub>	0.256 ***		0.262 ***		0.510 ***		0.291 **		0.015 *	
D <sub>xy</sub>	0.0051		0.0098		0.0117		0.0059		0.0059	
Tajima's D	−0.46 <sup>NS</sup>	−1.52 <sup>NS</sup>	−0.07 <sup>NS</sup>	+0.46 <sup>NS</sup>	−1.57 <sup>NS</sup>	−1.12 <sup>NS</sup>	−1.07 <sup>NS</sup>	−0.66 <sup>NS</sup>	−1.73 <sup>NS</sup>	−0.41 <sup>NS</sup>
Fu & Li's F	−0.19 <sup>NS</sup>	−1.89 <sup>NS</sup>	+0.07 <sup>NS</sup>	+0.34 <sup>NS</sup>	−2.31 *	−1.15 <sup>NS</sup>	−0.69 <sup>NS</sup>	−0.57 <sup>NS</sup>	−1.40 <sup>NS</sup>	+0.07 <sup>NS</sup>

N and S represent the number of sequences and the number of segregating sites used, respectively. Linkage disequilibria between sites were estimated between informative sites only: numbers in brackets correspond to the number of significant exact Fisher tests, total number of comparisons, and number of tests still significant after the Bonferroni correction, respectively. (RDP n.d.): Recombinant not detected using automated RDP and bootscan packages of RDP v.3.44. Values in brackets below  $\pi_s$  and  $\pi_n$  (Jukes & Cantor estimates) are the numbers of synonymous and nonsynonymous sites in coding regions, respectively. All genetic data sets obtained using the MCR method were corrected for artefactual/somatic singletons. \*:  $p = 0.05$ ; \*\*:  $p = 0.01$ ; \*\*\*:  $p = 0.001$ . <sup>NS</sup>: Non Significant. <sup>B</sup>: Bonferroni test.

### 3.3. Cryptic Amino Acid Variation along the AP-Pgm-1 Gene

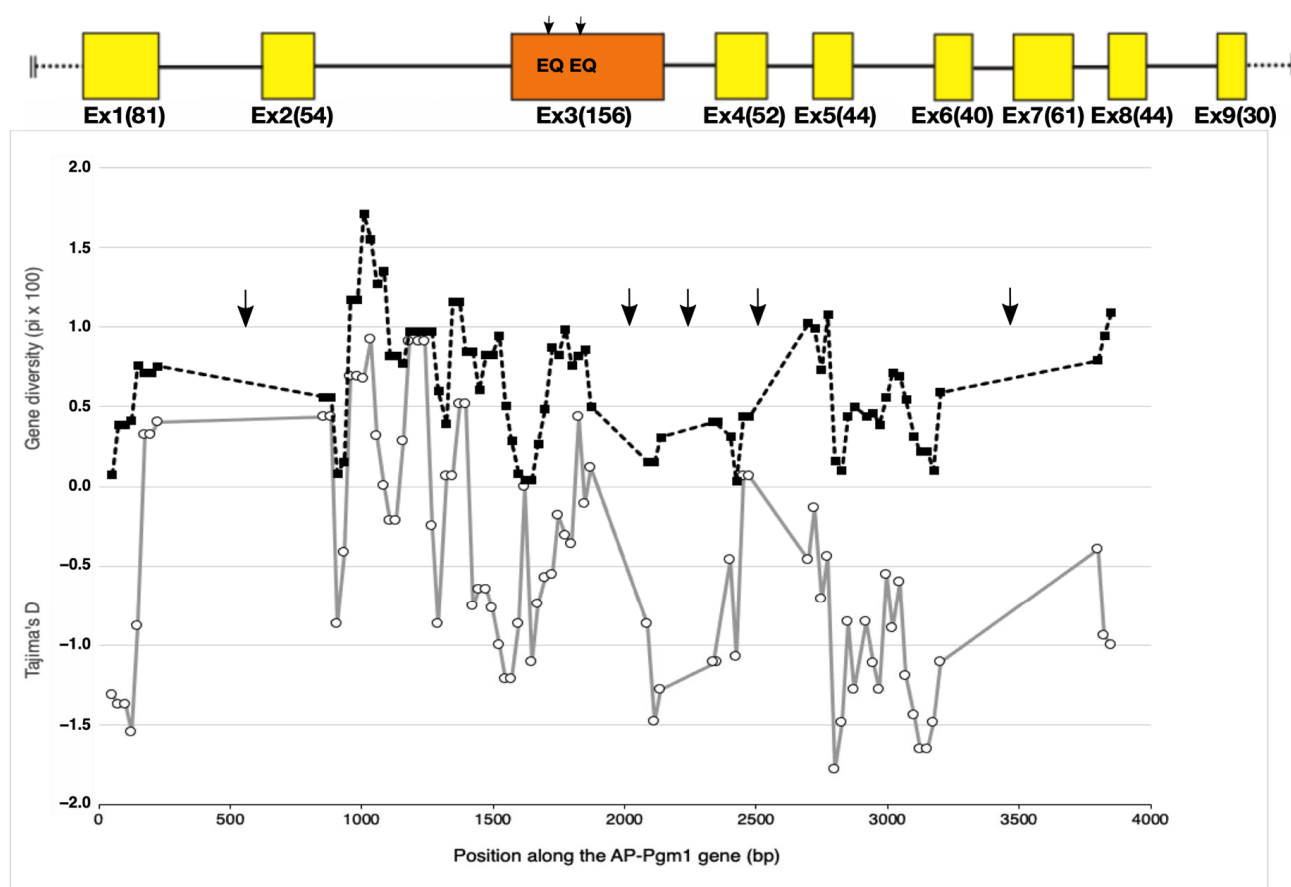
The full sequence of AP-Pgm-1 with the location of polymorphic codons and primers is shown in Figure S1. The total length of the nucleotidic sequence is 4372 bp. The gene is subdivided into nine exons and eight introns, with lengths ranging from 155 to 848 bp. The coding sequence of 1686 bp (562 codons) has an overall GC content of 43.5% (compared to only 29.3% in the intronic regions). When compared to human and oyster

*Pgm-1* genes [63,64], the largest *AP-Pgm-1* exon, comprising 156 codons (other exons vary from 40 to 81 codons), corresponds to the fusion of exons 3, 4, and 5 of the human *Pgm-1*. This fusion is shared with the oyster *C. gigas*, suggesting that annelids and molluscs share the same gene architecture (Figure 2).

Besides the two QE changes affecting the net charge of the protein in exon 3, other, less common, cryptic amino acid replacements were found along several regions of the *AP-Pgm-1* CDS. This allowed us to estimate gene diversities and the south/north divergence over an overall portion of about 3 kb (two-thirds of the gene; see Table 2). Gene diversities were almost constant over the *AP-Pgm-1* gene, but allele divergence increased locally in the vicinity of the two segregating EQ sites (Table 2). Looking more closely at the site variation along the gene using a sliding window on our set of sequenced fragments indicated that gene diversity was also slightly higher in exon 3, where the two QE substitutions were found with values almost identical to those depicted in intron 2 (Figure 3). This slight increase corresponded to peaks of positive Tajima's D values, which increased to +0.5 at the beginning of exon 3, suggesting that the presence of the two linked non-synonymous mutations may be associated with a gene diversity hotspot. However, the observed genetic diversities, as estimated from  $\theta_W$  and  $\pi$ , were not significantly greater than expected from neutral coalescent simulations for both the southern and northern populations over all the investigated *Pgm1* fragments (Figure 3, Table 2). Together with the two QE variant sites, the genotyping of exon 1 also confirmed the occurrence of a trans-equatorial V40L substitution, found at a frequency of 0.15 and restricted to the southern EQ allele (PGM90), and one of the two northern allelic lineages, regardless of the mutations EE (PGM100) and EQ (PGM90). The direct sequencing of the two other genic regions, located either between exons 4 and 5 and between exons 6 and 8, did not show any additional diagnostic amino acid changes between the three allelic lineages EE, EQ, and QE. In contrast, several synonymous changes and indels appeared to segregate between different allelic lineages along the gene (see sequence alignments, provided as supplementary data, for exons 1, 3, 4, 5, and 7 and introns 2, 6, and 7).

### 3.4. Estimating Allele Divergence and Linkage Disequilibria between Segregating Sites

To more specifically examine allele divergence and linkage disequilibria between segregating sites within allelic lineages, a sequencing of recaptured alleles was targeted on the longest region of the *AP-Pgm-1* gene (1110 bp). This region, containing intron 2 and the two allozyme-diagnostic sites EQ on exon 3, was genotyped from a subset of individuals. The sequencing of 48 alleles highlighted the presence of a high level of synonymous polymorphism with a strong linkage disequilibrium between these sites (Table 2), as well as two diagnostic indels in intron 2 (insertions referred to as A and B, following their order in the intron). These segregating sites and indels allowed us to determine four divergent allelic lineages with a few recombinants between them. These alleles were split between the northern and southern populations. In the southern population, the two allelic lineages L1 and L2 refer to the allozyme-diagnostic double mutation QE and EQ, whereas allelic lineages L3 and L4 refer to EQ and a mixture of EQ and EE in the northern population (Figure 4). At least nine and eleven synonymous mutations were fixed in intron 2 between allelic lineages L1 and L2, on one hand, and L3 and L4, on the other hand, respectively.



**Figure 3.** Evolution of gene diversity ( $\pi$ ) (black dashed line with black squares) and the Tajima's D statistic (grey line with white circles) along the *AP-Pgm-1* gene using a sliding window of 100 bp size with a 25 bp step. The analysis includes exonic and intronic fragments for which the sequence polymorphism has been documented. Arrows indicate the portions of the gene for which no genetic data set are available. Black arrows on *AP-Pgm-1* exon 3 indicate the positions of the two EQ substitutions.

In the southern population, allele L1 was typified by no insertion (QE, no\_A, no\_B), when compared to allele L2 (EQ and A, B), with a strong linkage disequilibrium between nearly all segregating sites (no recombination, see Table 2). It is worth noting, however, that one individual sampled at 7°25S originated from the northern populations had a L3L4 signature.

In the northern population, the two divergent lineages L3 and L4 also display linked sites, with either the A indel (L3) or the B indel (L4), but these two lineages were not completely associated with the double mutations EE and EQ. Alleles EE were mostly found in one of the two lineages plus one recombinant between L3 and L4, suggesting that these two lineages have recombined once (Figure 4; Table 2). Alternatively, allele EE could derive from one of the two lineages.

To estimate the recombination rate along the gene, we examined the distribution of the *Rho* parameter  $4N_r$  with Phase 2.1.1 over a greater proportion of the gene (1–2860 bp), using segregating sites ( $n = 53$ ) shared between northern and southern individuals that were successfully sequenced for all exon–intron fragments of *AP-Pgm-1* ( $N = 20$ ). Results from the Phase -recomb and -hotspot outputs clearly indicated that the recombination rate between alleles remained extremely low all along the gene (average local *Rho* = 0.033 and 0.008 for the northern and southern populations, respectively, which increased to nearly 2 in the southern population at the end of the gene, near the position 2140). Our results, therefore, indicated that the four allelic lineages greatly diverged from each other





was also close to zero. Introducing selection led to a better fit of the simulated parameters to the observed ones. Simulations with over-dominance and low recombination led to a slight decrease in  $F_{st}$  values (0.7) between demes, as well as an increase in the within-deme genetic diversities close to the observed ones, but also produced greater positive Tajima's  $D$  (+1.3 for each deme). Our best fit to values observed in the worm populations was obtained with the two-niche model simulations ( $F_{st} = 0.45$ ; nucleotide diversities  $\theta_W$  and Tajima's  $D$  estimates converging respectively, from 13 and +0.4 to 17 and +0.8 within and between demes). These simulated values were even closer to those estimated in the vicinity of the two EQ sites (intron 2; see Table 2 and Table S3).

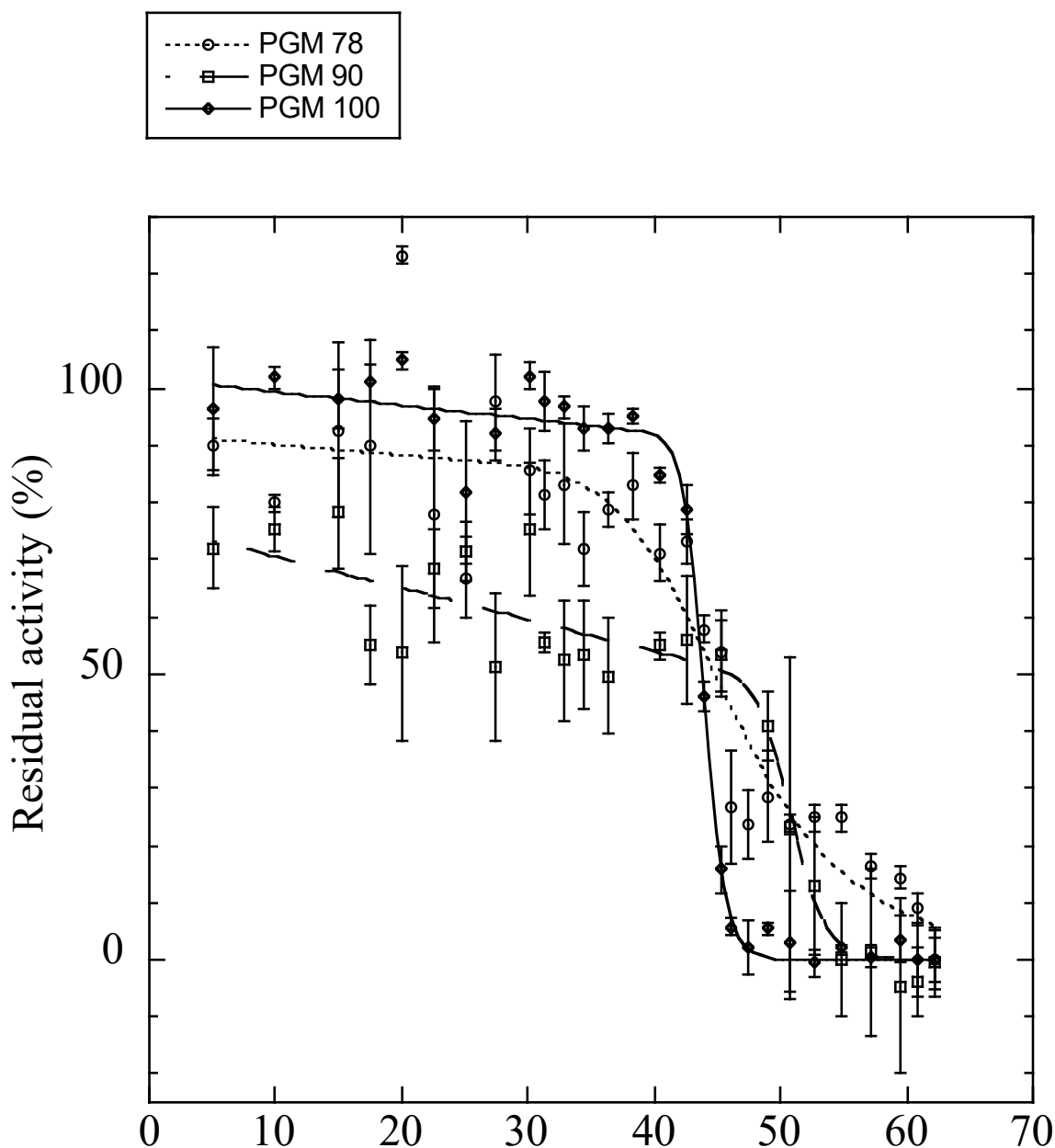
### 3.5. Conformational Stability, Thermal Inactivation, and Kinetics of the Mutated Isoforms

Obtaining full-length *AP-Pgm-1* cDNAs allowed us to examine the direct effect of the two QE substitutions on the thermal stability and efficiency of the PGM-1 enzyme using in vitro directed mutagenesis. To determine the conformational stability of the three recombinant isoforms of PGM-1, their guanidinium chloride (*GdmCl*)-induced unfolding was studied. Variations of fluorescence with increasing concentrations of *GdmCl* were biphasic (Figure S3), suggesting that the protein follows a three-state model of denaturation. For each transition, the unfolded fraction of protein ( $f_u$ ) was determined (Figure S4) and the Gibbs free energy change associated with each transition was calculated (Table 3). For the two transitions, PGM90 (EQ) appeared more stable than the two other isoforms. PGM78 (QE) appeared more stable than PGM100 for the first transition ( $\Delta G^0_{H_2O} = 8.0 \pm 0.46 \text{ kJ.mol}^{-1}$  vs.  $\Delta G^0_{H_2O} = 6.06 \pm 0.81 \text{ kJ.mol}^{-1}$ ), but not for the second transition ( $\Delta G^0_{H_2O} = 15.43 \pm 0.93 \text{ kJ.mol}^{-1}$  vs.  $\Delta G^0_{H_2O} = 15.13 \pm 0.98 \text{ kJ.mol}^{-1}$ ). The  $T_m$  values obtained from the theoretical curve fitted to the thermal inactivation experimental data (Figure 5, Table 3) were very similar for PGM78 ( $46.5 \pm 1.7 \text{ }^\circ\text{C}$ ) and PGM100 ( $44.0 \pm 0.1 \text{ }^\circ\text{C}$ ), but markedly higher for PGM90 ( $50.9 \pm 0.7 \text{ }^\circ\text{C}$ ).

**Table 3.** Conformational and temperature stability of the three expressed variants PGM78, PGM90, and PGM100.

	PGM 78		PGM 90		PGM 100	
	1st Transition	2nd Transition	1st Transition	2nd Transition	1st Transition	2nd Transition
$C_m$ (M)	$0.50 \pm 0.01$	$2.32 \pm 0.02$	$0.53 \pm 0.02$	$2.42 \pm 0.05$	$0.41 \pm 0.02$	$2.31 \pm 0.03$
$m$ ( $\text{kJ.mol}^{-1}.\text{M}^{-1}$ )	$16.00 \pm 0.88$	$6.65 \pm 0.39$	$21.62 \pm 3.73$	$7.48 \pm 1.13$	$10.45 \pm 1.28$	$6.55 \pm 0.42$
$\Delta G^0_{H_2O}$ ( $\text{kJ.mol}^{-1}$ )	$8.00 \pm 0.46$	$15.43 \pm 0.93$	$11.46 \pm 2.03$	$18.10 \pm 2.76$	$6.06 \pm 0.27$	$15.13 \pm 0.98$
$T_m$ ( $^\circ\text{C}$ )	$46.5 \pm 1.7$		$50.9 \pm 0.7$		$44 \pm 0.1$	
$K^{app}_m$ (mM)	$0.76 \pm 0.07$		$6.25 \pm 0.35$		$5.22 \pm 0.74$	
$K_{cat}$ ( $\text{sec}^{-1}$ )	$192 \pm 3.3$		$12.7 \pm 0.5$		$18.3 \pm 0.2$	
$K_{cat}/K^{app}_m$ ( $\text{sec}^{-1}.\text{M}^{-1}$ )	$252.63 \pm 17.06$		$2.0 \pm 0.1$		$3.51 \pm 0.49$	

$C_m$  and  $m$  values estimated from the variation of protein fluorescence in the presence of an increasing concentration of *GdmHCl* (values for each of the two transitions). Estimation of the free enthalpy of the unfolding reaction in the absence of the chaotropic agent for each of the two transition states.  $T_m$ : values of the temperature at which we reach 50% of non-reversible inactivation after a 30-min exposure.  $K^{app}_m$  and  $K_{cat}$  are kinetic parameters corresponding to the apparent Michaelis-Menten constant for glucose-1-phosphate and the catalytic constant, respectively. The ratio of these two values corresponds to the specific activity.



**Figure 5.** Residual enzyme activities after 30 min of incubation at different temperatures (on x-axis) for the expressed isoforms of PGM 78 (QE: dotted line with white circles), 90 (EQ: dashed line with white squares), and 100 (EE: plain line with black diamonds).  $T_m$  values are provided in Table 3.

Enzyme kinetic analyses of the three PGM isoforms are also presented in Table 3. The catalytic efficiency of PGM78, evaluated by the ratio  $k_{cat}/K_m^{app}$ , appeared to be 125 and 70-fold higher than that of PGM90 and PGM100, respectively. Changes in both  $K_m^{app}$  (for the substrate Glucose 1 phosphate, G1P) and  $k_{cat}$  explained most of the difference in the catalytic efficiency of the three isoforms.  $K_m^{app}$  (G1P) and  $k_{cat}$  of PGM78 were, indeed, tenfold lower and tenfold higher than that of the two other isoforms, respectively (see Table 3).

### 3.6. Fitness Cost of Individuals Carrying the Thermostable Allele in Terms of Female Fecundity

The Pompeii worm females exhibited an average coelomic fecundity of 200,000 oocytes, with great variability among them (values ranged from 1200 to 450,000 oocytes, depending on size, age, and reproductive status [42]). As opposed to our expectations, females carrying the allele 90 were, on average, more fecund than homozygous females carrying alleles 78

and 100; however, distributions of fecundity corrected by the size of the female did not significantly differ between *Pgm-1* genotypes (One-way ANOVA:  $F = 1.08$ ,  $p = 0.37$ ; see Figure S5). This finding clearly indicates that the ability to live under hotter conditions is not counter balanced by lower reproductive success; at least, for the females.

#### 4. Discussion

Based on allozyme data, Piccino et al. [12] previously proposed that the enzyme polymorphism of the Pompeii worm *A. pompejana* may be balanced at the locus *Pgm-1*, at least in populations of the northern EPR. They showed that allozymes 90 and 100, indeed, display distinct thermal stabilities and kinetic optima, with the frequency of the most thermostable isoform (allozyme 90) being positively correlated with temperature in newly formed edifices [12,36]. As the Pompeii worm is the only vent species capable of colonizing newly formed, still-hot hydrothermal chimneys, bearing thermostable alleles is likely to represent an adaptive advantage. The maintenance of polymorphism by selection on thermostable alleles, however, remains unresolved. If advantageous in the hottest part of the vent environment, thermostable alleles can be expected to spread rapidly in the population through recurrent selective sweeps. This is obviously not the case for the *Pgm-1* locus, which exhibited three major isoforms of different thermal stabilities (allozymes 78, 90, and 100), showing sharp frequency differences across the equatorial barrier to gene flow, as depicted by Plouviez et al. [40]. Several hypotheses have thus been proposed by Piccino et al. [12], regarding the maintenance of alleles at the PGM-1 enzyme. These includes: (i) allele over-dominance due to the rapid alternation of aerobic/anaerobic vent conditions; (ii) a fitness cost for individuals carrying the most thermostable allele at this locus; and (iii) a two-niche model effect, due to fluctuating proportions of ‘hot’ and ‘cold’ habitats along the EPR.

In the present study, we sequenced the three major *Pgm-1* alleles of the worm to investigate the distribution of non-synonymous polymorphisms along the gene, examine their relationship with allozymes, and assess their evolutionary fate. We demonstrated that only two linked mutations (E<sub>155</sub>Q and E<sub>190</sub>Q) are associated with the net charge of allozymes and are also responsible for the thermal performance of the three allozymes (78, 90, and 100). Examining the evolutionary history of these alleles indicated their rather old origin, which predates the vicariant event that separated the EPR vent fauna across the Equator about 1.2 million years ago [40,65,66]. In the following discussion, we therefore examine arguments towards the adaptive maintenance of PGM-1 isoforms. We propose that thermal compensation represents a powerful mechanism by which different enzymatic properties might be maintained under balancing selection—at least, during the exploration of the mutational landscape of the protein that will lead to the emergence of the ‘optimal’ isoform—in order to optimize metabolic fluxes, as previously stated by Eanes [67].

##### 4.1. Two-Allele Polymorphism at the *Pgm* Locus: A Long Story of Balancing Selection?

The non-synonymous polymorphisms associated with the four allelic lineages of *AP-Pgm-1* appeared to be quite low ( $\pi_N = 0.0025$ , on average). Such a result sharply contrasts with earlier studies on branch-point glycolytic enzymes (that control the metabolic flux for transport, storage, and breakdown of carbohydrates), for which numerous cryptic non-synonymous changes were described [67]. High levels of gene diversity have indeed been observed between the slow, medium, and fast electrophoretic *Pgm-1* alleles of *Drosophila melanogaster* [14,68], and between phosphoglucose isomerase (PGI) alleles of *Colias* butterflies [69], suspected to evolve under balancing selection. By contrast, only eight nonsynonymous mutations (E37Q, V40L, E155Q, E190Q, R343I, G358S, T366M, and F502L), some of which were at low frequency, were detected by the direct comparison of allelic sequences of *A. pompejana*. Moreover, even if gene diversity was slightly higher in the intronic region preceding exon 3 and the exon 3 itself where the two EQ sites are found, its variation along the gene does not fit perfectly with the expectations of long-term balancing selection. A weak hotspot signal of silent site variation, supported by slightly positive Tajima’s D and

Fu & Li's  $F$  statistics was, however, observed near the doubly selected sites  $E^{155}Q/E^{190}Q$ , but not as strong as signals depicted for the *Adh* locus in *Drosophila* [70,71]. Theoretical effects of balancing selection on nearby genome regions promote increase in genetic diversity near the selected site due to lack of recombination and the long-term accumulation of mutations [38]. The very low level of nucleotidic polymorphism found at *AP-Pgm-1* can, however, be partially explained by recurrent population bottlenecks, due to the challenging environmental conditions that affect the whole vent fauna [15]. The joint action of abrupt demographic changes and habitat specialization should, indeed, promote enzyme monomorphism. Under such conditions, the level of polymorphism observed at *Pgm-1*, although low, appears to be quite unusual when compared to most of the genes examined in *A. pompejana*. In alvinellid worms—and, especially, thermophilic species—proteins are under strong purifying selection, with transcriptome-wide  $d_N/d_S$  averages very close to zero, and individual gene values ranging between 0.02 and 0.05 [72]. To this extent, it is worth noting that gene diversity at the *Pgm-1* locus appears to be locally two to four-fold higher than the genome-wide average value (ddRAD overall  $\pi = 0.0025$  [73]), and from other reported genes [35].

Looking more specifically at the four allelic lineages of *AP-Pgm1* near the EQ sites (intron 2) clearly indicates that they have accumulated a great number of synonymous substitutions due to their separation with almost no recombination events (low values of  $R_m$  and  $R_{ho}$ , see the Phase 2.1.1 analysis). The two allelic lineages, L1 and L2, present in the southern population exhibited 1% divergence between them, with a strong linkage disequilibrium between the two variant sites  $E^{155}Q^{190}$  (PGM90) and  $Q^{155}E^{190}$  (PGM78), and the silent substitutions found in intron 2. This suggests that the two allelic lineages evolved separately, without recombination in the vicinity of the two nonsynonymous sites, for a long period of time. The two northern allelic lineages (L3 and L4) also diverged by 0.7% and display two diagnostic indels in intron 2. These silent mutations and indels are also linked, suggesting once more that these two lineages evolved separately, but for a shorter period of time. However, diagnostic mutations were not completely linked to the variant sites  $E^{155}E^{190}$  (PGM100) and  $E^{155}Q^{190}$  (PGM90). Although L3 forms a single clade associated with  $E^{155}Q^{190}$ , L4 is a mixture of  $E^{155}E^{190}$  and  $E^{155}Q^{190}$ , suggesting either that at least one recombination event occurred between the two northern alleles, or that  $E^{155}E^{190}$  (PGM100) is a recently derived variant in the northern population. Finally, both divergences observed between alleles within each population were of the same amplitude as the divergence estimated between the southern and northern alleles (0.9%), which coincided with an overall significant  $F_{st}$  value of 0.588 between them. If we accept that the southern (L1 and L2) and northern (L3 and L4) allelic lineages became isolated after the appearance of the physical barrier to dispersal—about 1.2 million years ago [35,40]—this clearly indicates that the co-occurrence of the four highly divergent *Pgm-1* alleles derives from an older polymorphism, predating the vicariant event that separated the northern and southern vent fauna of the East Pacific Rise, with the possible emergence of the genotype  $E^{155}E^{190}$  in the north. Such a scenario is likely confirmed by the distribution of alleles in the exon 3 haplotype network and the signature of the linked silent polymorphic sites in introns 1 and 2, where northern alleles seem to derive from a southern allele bearing the  $Q^{155}E^{190}$  mutation.

#### 4.2. Selective Modalities for the Maintenance of a Balanced Polymorphism

The long-term evolution of *Pgm-1* alleles without recombination—at least, in the first part of the gene—and their frequency changes according to environmental conditions, raises questions regarding the selective modalities acting on the co-occurrence of alleles when one of the two alleles is better adapted to high temperatures [12]. One of the first explanations for the adaptive maintenance of AP-PGM1 isoforms was to consider that the rapid alternation of oxic and anoxic conditions during venting should favour heterozygote excesses if the heterozygote's fitness is close to that of the favoured homozygote in one of the two habitats [74]. Pogson [13] previously proposed that over-dominance represents



the most likely evolutionary mechanism at the origin of the maintenance of a balanced polymorphism at the *Pgm-2* locus for the oyster *C. gigas* (see also [75], for its link with the individual's growth rate). Here, we were not able to detect any overdominance at the *Pgm-1* locus in terms of heterozygote excesses in natural populations. Simulations of structured coalescent with asymmetrical migration and over-dominance always led to very high within-deme positive Tajima's D and unequal diversities between overall and within demes (not found in the observed dataset). This finding confirms a previous study [12] considering allozymes. However, it is worth noting that such an advantage can be easily masked by the temporal dynamics of the thermal habitat (i.e., chimneys refreshing with time) and the juxtaposition of chimneys of different ages. The worm is, indeed, exposed to a mosaic of fluctuating thermal habitats where temperature could vary spatially, according to the age of the chimneys [12].

Maintenance of allozymes with different thermal stabilities can be also explained by a two-niche model of local differentiation between habitats and drift [76–79]. Simulating coalescences with a two-niche model and a theta value equal to the observed value provided parameter estimates ( $F_{st}$ , as well as diversities and Tajima's D) much closer to values observed in the vicinity of the two EQ sites than the two other models of asymmetric gene flow with and without over-dominance. Given the spatial and temporal dynamics of the hydrothermal discharge, changes in the frequency of *Pgm-1* alleles could be either due to local selection or exacerbated genetic drift associated with the dynamics of colonization of the newly opened sites. Indeed, the dynamic nature of hydrothermal vents over longer time scales (i.e., years) led to a very patchy and transient habitat scattered along the EPR, with a complex heterogeneity of age-driven vent conditions. This can be seen as a multitude of distinct ecological niches for the same species. In this context, the proportion of newly formed 'still hot' chimneys and older 'colder' ones greatly vary over time. These dynamics depends on the spreading rate of the rift and, thus, the frequency of tectonic and volcanic events along the East Pacific Rise.

To explain the maintenance of a bi-allelic polymorphism at the *Pgm-1* locus, Piccino et al. [12] proposed a fitness cost of worm colonization during the early stages of a chimney. The first settlers on 'hot' (>100 °C) anhydrite chimneys benefit from a lack of predators and competitors. Colonists may, therefore, display more thermoresistant alleles but lower reproductive investment and/or survival. Watt et al. [80–82] and Wheat et al. [69] previously showed that both PGI and PGM have a great contribution to the mating fitness of male *Colias* butterflies, likely as the result of longer and more vigorous flight within the day. However, based on female fecundity, we were not able to observe fitness differences between the *Pgm-1* genotypes. This suggests that a better predisposition to colonize still 'hot' chimneys is probably not compensated by reduced reproductive success to prevent fixation of the advantageous allele. In fruit flies, the *Pgm* locus represents a quantitative trait for glycogen storage and, hence, the ability to survive better under starvation [34]. In the case of *A. pompejana*, differences in the thermal regime could be great between colonists and reproducers. As a consequence, colonists subjected to longer periods of high temperature (and associated hypoxia) may be maladapted to produce and use their glycogen reserves. This should have consequences in their investment into reproduction. To the contrary, secondary settlers arriving in much cooler conditions are more likely to use their glycogen reserves to massively invest in the production of gametes, as was previously shown by Faure et al. [42]. Colder conditions seem to be a prerequisite for releasing fertilized eggs after pairing, as embryos are not able to develop at temperatures greater than 15 °C [41].

#### 4.3. Adaptive Polymorphism: A Trade-Off between Enzyme Thermostability and Catalysis

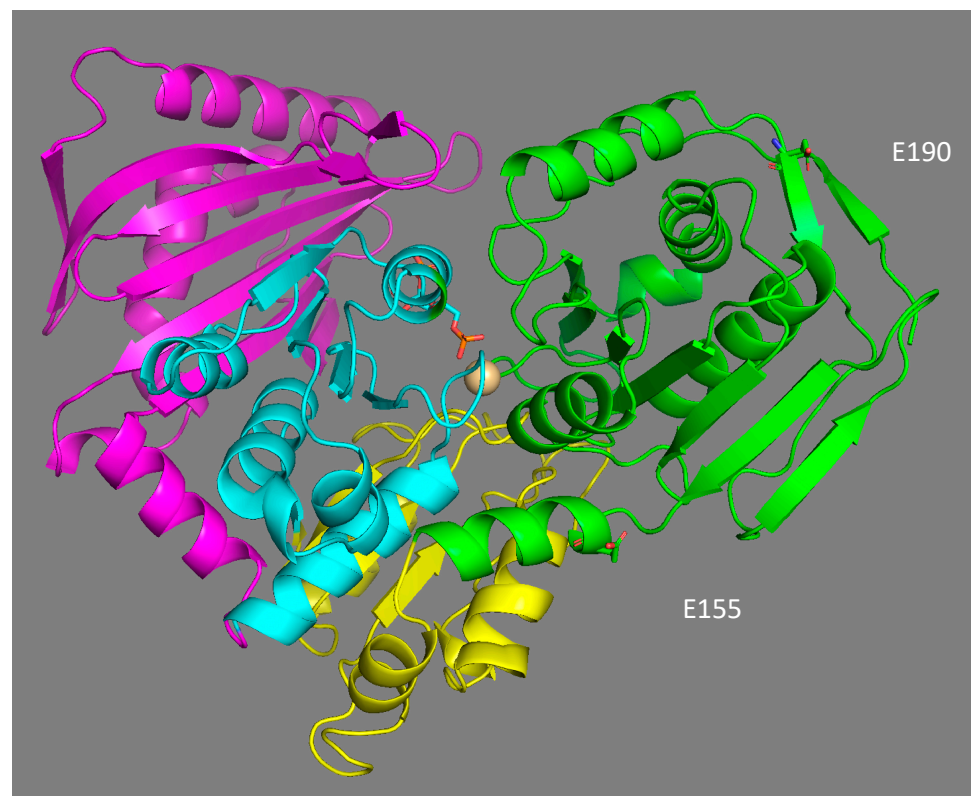
In *Drosophila*, *Pgm-1* variants play a non-negligible role in regulating the metabolic energy pool along latitudinal clines, where a decrease in temperature is compensated for by an increase in enzyme activity. Populations of *D. melanogaster* living at the highest (and, thus, coldest) latitudes possess PGM allozymes with a higher catalytic efficiency and greater glycogen contents. According to the theory of metabolic flux, this can be an adaptive

means of temperature compensation to maintain the same glycogen contents over the latitudinal gradient [67]. Differences in both protein thermostability and catalytic efficiency between *Pgm* alleles have been previously reported to explain both local differentiation and latitudinal clines in the oyster *C. gigas* [6,13] and *D. melanogaster* [34]. By comparison, thermal compensation may be directly linked to different biochemical phenotypes that interact with the growth rate and reproductive effort of the worms. The theory predicts that differences in activity at only one enzyme must be substantial to affect metabolic fluxes between genotypes [83]. To test this hypothesis, one could measure the effect of non-synonymous mutations on the functional properties of the enzyme, its conformational stability, and their effects on population fitness. In this study, three recombinant isoforms of PGM-1—E<sup>155</sup>E<sup>190</sup> (PGM100), E<sup>155</sup>Q<sup>190</sup> (PGM90), and Q<sup>155</sup>E<sup>190</sup> (PGM78)—were obtained by directed mutagenesis. The replacement of the glutamate by a glutamine at position 190 increased the conformational stability and thermostability of the protein, confirming that PGM90 is the most thermostable isoform. As discussed by Piccino et al. [12], carrying this allele may be advantageous during the colonization of newly formed chimneys, whose surface temperature usually exceeds 50 °C. Unexpectedly, PGM90 exhibited a decrease in the catalytic efficiency of the enzyme when compared to the two other recombined variants ( $k_{cat}/K_m$  was a hundred times for PGM78 and nearly two-fold greater for PGM100, when compared to PGM90). The recombinant PGM90 also exhibited the lowest affinity for its substrate, glucose-1-phosphate, at 17 °C. This finding is of importance, as such a genetically determined trade-off between protein stability and enzyme activity has not, so far, been reported in other invertebrate species subjected to balancing selection [6,13,34,82]. Increased thermostability of a protein is often associated with a decrease in the flexibility of the molecule and, thus, the dynamics of the enzyme reaction [84,85]. Our results are in perfect agreement with these theoretical expectations. They support the positive role of a thermodynamic trade-off between thermostability and catalysis, as previously proposed by Eanes [11], to explain the co-occurrence of alleles. PGM90 can remain stable for a longer period of time but is less efficient in either producing or consuming the glycogen reserves of the worm than the two other isoforms. To this extent, the fact that the  $k_{cat}/K_m$  ratio of isoform PGM78 is much higher than that of the isoform PGM100 can explain why isoform 78 is more frequent in the southern populations (about 80%), when compared with isoform PGM100 in the northern populations (around 70%). The balance between allele frequencies from both sides of the Equator may be dictated by the selective coefficient attributed to each genotype as a direct reflection of the catalytic efficiency difference between PGM90 and its alternative isoforms.

#### 4.4. Structural Effect of Mutations E155Q and E190Q

The location of the two main polymorphic sites (E155Q and E190Q) in the 3D model structure of phosphoglucomutase 1 expose them to solvents, and are not located in the binding domains of the enzyme (Figure 6). Their potential effect on the catalytic properties of the enzyme is, therefore, not the result of a direct interaction with the substrate and/or the residues involved in catalysis. This is not surprising, as most of the mutations affecting the binding of the substrate glucose-1-phosphate, the Mg<sup>2+</sup> ion, and phosphate should be deleterious. Similarly, in a study of the polymorphism of the enzyme PGM in *D. melanogaster*, none of the 21 polymorphic amino acid replacements were located in the catalytic site of the enzyme [34]. Based on their location, both substitutions should affect the net charge of the protein in the same way. However, the isoforms 78 and 90 did not have the same electrophoretic mobility, suggesting that some post-translational modification may be involved in the electrophoretic separation of the three isoforms. The gain of a glutamate at position 155 may be associated with a potential ionic bond with histidine 157, with a distance of 6.5 Å between them. Ionic and hydrogen bonds have been shown to increase the stability of enzymes [86] and partially explain the thermostable 3D structure of the Cu–Zn and Mn superoxide dismutase enzymes in *A. pompejana* [87,88]. This may account for the increased thermostability of isoform 90 but should also have the same effect for isoform

100, which was obviously not the case. This suggests a negative effect of glutamate (E) at position 190, which negates the positive effect of glutamate at position 155. Alternatively, the 3D model comparison of the three isoforms shows that the replacement of one glutamine (Q) by a glutamate (E) at position 190 (as in allozymes 78 and 100) introduces a negative charge in a region already enriched in acidic residues. This high density of negative charges could have a destabilizing effect on the protein structure through a Coulomb repulsion effect, thus potentially leading to greater sensitivity to temperature. Finally, the glutamine replacement at position 155 is also likely to play a key role in the molecular dynamics of the protein, especially during the 180° rotation of the reaction intermediate (glucose-1,6-diphosphate) inside the active site. This potentially explains the higher enzymatic efficiency of the isoform 78.



**Figure 6.** 3-D structural model of *A. pompejana* PGM 78 fitted to the PGM-1 rabbit template (1C47, 2.70 Å) using Modeller 9v13 software. The protein is structured in four domains, labelled from I to IV (I green, II yellow, III blue, IV violet). Positions 155 and 190 of EQ replacements belong to domain I near the catalytic site of the enzyme, which binds the reaction catalyser,  $\alpha$ -D-glucose-1,6-diphosphate, and the  $Mg^{2+}$  ion.

## 5. Conclusions

Exploration of the mutational landscape of phosphoglucomutase 1 revealed the maintenance of four divergent allelic lineages over the geographic range of *A. pompejana*. The enzyme polymorphism, only governed by two linked amino acid replacements (E155Q and E190Q), is likely to be maintained by balancing selection, and better fits with a two-niche model of selection, in which ‘cold’ and ‘hot’ conditions alternate. These isoforms do not seem to be maintained by higher reproductive fitness of females that do not carry the thermostable allele needed during the colonisation of new ‘hot’ chimneys. This persistence over a long period of time may rather be explained by a thermodynamic trade-off between protein thermostability and catalysis of these functional phenotypes.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/genes13020206/s1>. Figure S1: Entire sequence of the *Pgm-1* gene of *A. pompejana* and its translated exons. Figure S2: Haplotype network obtained by genotyping of the *Pgm-1* exon 3 on 187 individuals coming from the northern and the southern EPR. Figure S3: Guanidium chloride (*GdmCl*) denaturation curves for the three PGM1 expressed isoforms. Figure S4: Modelled regression curves of the folded/unfolded protein states  $f_u(I)$  and  $f_u(II)$  of three PGM1 expressed isoforms for each transition according to the denaturation equilibrium  $N \leftrightarrow I \leftrightarrow U$  (Native  $\leftrightarrow$  Intermediate  $\leftrightarrow$  Unfolded). Figure S5: Scatterplots of female *A. pompejana* fecundities according to their PGM-1 allozyme genotypes. Table S1: Sequence of the primer pairs, and their use in the study. Table S2: Frequencies of EE, EQ and QE *Pgm-1* alleles, heterozygosities and Fis (\*: Significant with 1000 permutations) in northern and southern populations of *A. pompejana*. Table S3: Summary statistics of structured coalescent simulations ( $n = 1000$ ) obtained with MSMS software and the pylibseq libraries in order to test a model of asymmetric migration across a barrier with and without selection and two levels of recombination in order to examine the genetic expectations of overdominance and the two-niches (two demes/two habitats) models.

**Author Contributions:** Conceptualization, D.J.; formal analyses, A.B., E.S., M.V., B.F., J.M., P.P. and D.J.; methodology, A.-S.L.P.; sampling of resources, B.F., P.P., S.H. and D.J.; data curation, D.J.; writing—original draft preparation, A.B., J.M. and D.J.; writing—review and editing, S.H. and D.J.; supervision, D.J.; funding acquisition, S.H. and D.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** The present study was supported by TOTAL Foundation and the CNRS.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This article was written in memory of Dominique Le Guen, who greatly helped us in setting up of the allozyme genotyping. We thank chief scientists and the ‘Nautile’ crews for their technical support and efforts during the oceanographic expeditions Phare2002, Biospeedo2004, and Mescal2010. We are very grateful to the two anonymous referees who provided valuable comments and editorial suggestions on the manuscript, and to Pierre Alexandre Gagnaire and Thomas Broquet who edited the manuscript to improve its readability. We are also indebted to K. Thornton who wrote a python script to use his pylibseq libraries and parse the output of the MSMS program.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Holt, R.D.; Gaines, M.S. Analysis of adaptation in heterogeneous landscapes: Implications for the evolution of fundamental niches. *Evolut. Ecol.* **1992**, *6*, 433–447. [[CrossRef](#)]
- Schmidt, P.S.; Serrão, E.A.; Pearson, G.A.; Riginos, C.; Rawson, P.D.; Hilbish, T.J.; Brawley, S.H.; Trussel, G.C.; Carrington, E.; Wetthey, D.S.; et al. Ecological genetics in the North Atlantic: Environmental gradients and adaptation at specific loci. *Ecology* **2008**, *89*, 91–107. [[CrossRef](#)]
- Nevo, E.; Shimony, T.; Libni, M. Thermal selection of allozyme polymorphisms in barnacles. *Nature* **1977**, *267*, 99. [[CrossRef](#)]
- Nevo, E.; Lavie, E.; Ben-Shlomo, R. Selection of allelic isozyme polymorphisms in marine organisms: Pattern, theory, and application. *Isozymes* **1983**, *10*, 69–92. [[PubMed](#)]
- Nevo, E.; Noy, R.; Lavie, B.; Beiles, A.; Muchtar, S. Genetic diversity and resistance to marine pollution. *Biol. J. Linn. Soc.* **1986**, *29*, 39–144. [[CrossRef](#)]
- Pogson, G.H. Biochemical characterization of genotypes at the phosphoglucosyltransferase-2 locus in the Pacific oyster, *Crassostrea gigas*. *Biochem. Genet.* **1989**, *27*, 571–589. [[CrossRef](#)]
- Riddoch, B. The adaptive significance of electrophoretic mobility in phosphoglucose isomerase (PGI). *Biol. J. Linn. Soc.* **1993**, *50*, 1–17. [[CrossRef](#)]
- Schmidt, P.S.; Rand, D.M. Adaptive maintenance of genetic polymorphism in an intertidal barnacle: Habitat- and life-stage-specific survivorship of MPI genotypes. *Evolution* **2001**, *55*, 1336–1344. [[CrossRef](#)]
- Boutet, I.; Tanguy, A.; le Guen, D.; Piccino, P.; Hourdez, S.; Ravaux, J.; Shillito, B.; Legendre, P.; Jollivet, D. Global depression in gene expression as a response to rapid thermal changes in vent mussels. *Proc. R. Soc. Lond. B* **2009**, *276*, 3071–3079. [[CrossRef](#)] [[PubMed](#)]



10. Bougerol, M.; Boutet, I.; le Guen, D.; Jollivet, D.; Tanguy, A. Transcriptomic response of the hydrothermal mussel *Bathymodiolus azoricus* mediated by heavy metals is modulated by *Pgm* genotypes and symbiont content. *Mar. Genom.* **2015**. [[CrossRef](#)]
11. Eanes, W.F. Analysis of selection on enzyme polymorphisms. *Ann. Rev. Ecol. Syst.* **1999**, *30*, 301–326. [[CrossRef](#)]
12. Piccino, P.; Viard, F.; Sarradin, P.-M.; le Bris, N.; le Guen, D.; Jollivet, D. Thermal selection of PGM allozymes in newly founded populations of the thermotolerant vent polychaete *Alvinella pompejana*. *Proc. R. Soc. Lond. B* **2004**, *271*, 2351–2359. [[CrossRef](#)] [[PubMed](#)]
13. Pogson, G.H. Expression of overdominance for specific activity at the phosphoglucosyltransferase-2 locus in the Pacific oyster, *Crassostrea gigas*. *Genetics* **1991**, *128*, 133–141. [[CrossRef](#)] [[PubMed](#)]
14. Verrelli, B.C.; Eanes, W.F. Clinal variation for amino acid polymorphisms at the *Pgm* locus in *Drosophila melanogaster*. *Genetics* **2001**, *157*, 1649–1663. [[CrossRef](#)]
15. Vrijenhoek, R.C. Gene flow and genetic diversity in naturally fragmented metapopulations of deep-sea hydrothermal vent animals. *J. Hered.* **1997**, *88*, 285–293. [[CrossRef](#)]
16. Jollivet, D.; Chevalloné, P.; Planque, B. Hydrothermal-vent alvinellid polychaete dispersal in the Eastern Pacific. 2. A metapopulation model based on habitat shifts. *Evolution* **1999**, *53*, 1128–1142. [[CrossRef](#)]
17. Vrijenhoek, R.C. Genetic diversity and connectivity of deep-sea hydrothermal vent metapopulations. *Mol. Ecol.* **2010**, *19*, 4391–4411. [[CrossRef](#)]
18. Watremez, P.; Kerveyan, C. 1990 Origine des variations de l'activité hydrothermale: Premiers éléments de réponse d'un modèle numérique simple. *C. R. Acad. Sci Paris* **2010**, *311*, 153–158.
19. Johnson, K.S.; Childress, J.J.; Hessler, R.R.; Sakamoto-Arnold, C.M.; Beehler, C.L. Chemical and biological interactions in the Rose Garden hydrothermal vent field, Galapagos spreading center. *Deep Sea Res. A* **1988**, *35*, 1723–1744. [[CrossRef](#)]
20. Le Bris, N.; Gaill, F. How does the annelid *Alvinella pompejana* deal with an extreme hydrothermal environment? *Rev. Environ. Sci. Biotechnol.* **2007**, *6*, 197–221. [[CrossRef](#)]
21. Hourdez, S.; Lallier, F.H. Adaptations to hypoxia in hydrothermal-vent and cold-seep invertebrates. In *Life in Extreme Environments*; Springer: Dordrecht, The Netherlands, 2006; pp. 297–313.
22. Von Damm, K.L. Controls on the chemistry and temporal variability of seafloor hydrothermal fluids. In *Seafloor Hydrothermal Systems: Physical, Chemical, Biological, and Geological Interactions*; Washington DC American Geophysical Union Geophysical Monograph Series; Wiley: Hoboken, NJ, USA, 1995; Volume 91, pp. 222–247.
23. Jollivet, D. Specific and genetic diversity at deep-sea hydrothermal vents: An overview. *Biodiv. Cons.* **1996**, *5*, 1619–1653. [[CrossRef](#)]
24. Matabos, M.; le Bris, N.; Pendlebury, S.; Thiébaud, E. Role of physico-chemical environment on gastropod assemblages at hydrothermal vents on the East Pacific Rise (13 N/EPR). *J. Mar. Biol. Assoc. UK* **2008**, *88*, 995–1008. [[CrossRef](#)]
25. Shank, T.M.; Fornari, D.J.; von Damm, K.L.; Lilley, M.D.; Haymon, R.M.; Lutz, R.A. Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9 50' N. East Pacific Rise). *Deep Sea Res.* **1998**, *45*, 465–515. [[CrossRef](#)]
26. Fontaine, F.J.; Cannat, M.; Escartin, J. Hydrothermal circulation at slow-spreading mid-ocean ridges: The role of along-axis variations in axial lithospheric thickness. *Geology* **2008**, *36*, 759–762. [[CrossRef](#)]
27. Marcus, J.; Tunncliffe, V.; Butterfield, D.A. Post-eruption succession of macrofaunal communities at diffuse flow hydrothermal vents on Axial Volcano, Juan de Fuca Ridge, Northeast Pacific. *Deep Sea Res.* **2009**, *56*, 1586–1598. [[CrossRef](#)]
28. Chevalloné, P.; Desbruyères, D.; Childress, J.J. ... and some even hotter. *Nature* **1992**, *359*, 593–594. [[CrossRef](#)]
29. Cary, S.C.; Shank, T.; Stein, J. Worms bask in extreme temperatures. *Nature* **1998**, *391*, 545–546. [[CrossRef](#)]
30. Ravaux, J.; Hamel, G.; Zbinden, M.; Tasiemski, A.A.; Boutet, I.; Léger, N.; Tanguy, A.; Jollivet, D.; Shillito, B. Thermal limit for metazoan life in question: In Vivo heat tolerance of the Pompeii worm. *PLoS ONE* **2013**, *8*, e64074.
31. Jang, S.J.; Park, E.; Lee, W.K.; Johnson, S.B.; Vrijenhoek, R.C.; Won, Y.J. Population subdivision of hydrothermal vent polychaete *Alvinella pompejana* across equatorial and Easter Microplate boundaries. *BMC Evol. Biol.* **2016**, *16*, 1–15. [[CrossRef](#)]
32. Desbruyères, D.; Chevalloné, P.; Alayse, A.-M.; Jollivet, D.; Lallier, F.H.; Jouin-Toulmond, C.; Zal, F.; Sarradin, P.-M.; Cosson, R.; Caprais, J.-C.; et al. Biology and ecology of the “Pompeii worm” (*Alvinella pompejana* Desbruyères and Laubier), a normal dweller of an extreme deep-sea environment: A synthesis of current knowledge and recent developments. *Deep Sea Res.* **1998**, *45*, 383–422. [[CrossRef](#)]
33. Chevalloné, P.; Fisher, C.R.; Childress, J.J.; Desbruyères, D.; Jollivet, D.; Zal, F.; Toulmond, A. Thermotolerance and the ‘Pompeii worms’. *Mar. Ecol. Prog. Ser.* **2000**, *208*, 293–295. [[CrossRef](#)]
34. Jollivet, D.; Desbruyères, D.; Bonhomme, F.; Moraga, D. Genetic differentiation of deep-sea hydrothermal vent alvinellid populations (Annelida: Polychaeta) along the East Pacific Rise. *Heredity* **1995**, *74*, 376–391. [[CrossRef](#)]
35. Plouviez, S.; le Guen, D.; Lecompte, O.; Lallier, F.H.; Jollivet, D. Determining gene flow and the influence of selection across the equatorial barrier of the East Pacific Rise in the tube-dwelling polychaete *Alvinella pompejana*. *BMC Evol. Biol.* **2010**, *10*, 220. [[CrossRef](#)] [[PubMed](#)]
36. Jollivet, D.; Desbruyères, D.; Ladrat, C.; Laubier, L. Evidence for differences in the allozyme thermostability of deep-sea hydrothermal vent polychaetes (Alvinellidae): A possible selection by habitat. *Mar. Ecol. Prog. Ser.* **1995**, *123*, 125–136. [[CrossRef](#)]
37. Verrelli, B.C.; Eanes, W.F. The functional impact of *Pgm* amino acid polymorphism on glycogen content in *Drosophila melanogaster*. *Genetics* **2001**, *159*, 201–210. [[CrossRef](#)]



38. Charlesworth, D. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* **2006**, *2*, e64. [[CrossRef](#)] [[PubMed](#)]
39. Hedrick, P.W. Balancing selection. *Curr. Biol.* **2007**, *17*, 230–231. [[CrossRef](#)]
40. Plouviez, S.; Shank, T.M.; Faure, B.; Daguin-Thiébaud, C.; Viard, F.; Lallier, F.H.; Jollivet, D. Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South. *Mol. Ecol.* **2009**, *18*, 3903–3917. [[PubMed](#)]
41. Pradillon, F.; Shillito, B.; Young, C.M.; Gaill, F. Deep-sea ecology: Developmental arrest in vent worm embryos. *Nature* **2001**, *413*, 698. [[CrossRef](#)]
42. Faure, B.; Chevaldonné, F.P.; Jollivet, D. Spatial and temporal dynamics of reproduction and settlement in the Pompeii worm *Alvinella pompejana* (Polychaeta: Alvinellidae). *Mar. Ecol. Prog. Ser.* **2007**, *348*, 197–211. [[CrossRef](#)]
43. Mishra, R.N.; Singla-Pareek, S.L.; Nair, S.; Sopory, S.K.; Reddy, M.K. Directional genome walking using PCR. *Biotechniques* **2002**, *33*, 830–842. [[CrossRef](#)] [[PubMed](#)]
44. Bierne, N.; Tanguy, A.; Faure, M.; Faure, B.; David, E.; Boutet, I.; Boon, E.; Quéré, N.; Plouviez, S.; Kemppainen, P.; et al. Mark-recapture cloning: A straightforward and cost-effective cloning method for population genetics of single copy nuclear DNA sequences in diploids. *Mol. Ecol. Notes* **2007**, *7*, 562–566. [[CrossRef](#)]
45. Black, W.C.; Krafur, E.S. A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. *Theor. Appl. Genet.* **1985**, *70*, 491–496. [[CrossRef](#)] [[PubMed](#)]
46. Belkhir, K.; Borsa, P.; Chikhi, L.; Raufaste, N.; Bonhomme, F. *GENETIX 4.0. 5.2, Software under Windows™ for the Genetics of the Populations*; University of Montpellier: Montpellier, France, 2004.
47. Bandelt, H.J.; Forster, P.; Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **1999**, *16*, 37–48. [[CrossRef](#)]
48. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [[CrossRef](#)]
49. Kelly, J.K. A test of neutrality based on interlocus associations. *Genetics* **1997**, *146*, 1197–1206. [[CrossRef](#)]
50. Hudson, R.R.; Kaplan, N.L. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **1985**, *111*, 147–164. [[CrossRef](#)]
51. Martin, D.P.; Lemey, P.; Lott, M.; Moulton, V.; Posada, D.; Lefevre, P. RDP3: A flexible and fast computer program for analyzing recombination. *Bioinformatics* **2010**, *26*, 2462–2463. [[CrossRef](#)]
52. Li, N.; Stephens, M. Modelling linkage disequilibrium and identifying recombination hotspots using snp data. *Genetics* **2003**, *165*, 2213–2233. [[CrossRef](#)]
53. Hudson, R.R.; Slatkin, M.; Maddison, W.P. Estimation of levels of gene flow from DNA sequence data. *Genetics* **1992**, *132*, 583–589. [[CrossRef](#)] [[PubMed](#)]
54. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)] [[PubMed](#)]
55. Ewing, G.; Hermisson, J. MSMS: A coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* **2010**, *26*, 2064–2065. [[CrossRef](#)] [[PubMed](#)]
56. Thornton, K. Libsequence: A c++ class library for evolutionary genetic analysis. *Bioinformatics* **2003**, *19*, 2325–2327. [[CrossRef](#)]
57. Reikofski, J.; Tao, B.Y. Polymerase chain reaction (PCR) techniques for site-directed mutagenesis. *Biotechnol. Adv.* **1992**, *10*, 535–547. [[CrossRef](#)]
58. Pace, C.N.; Scholtz, J.M. Measuring the conformational stability of a protein. In *Protein Structure: A Practical Approach*; Oxford University Press: Oxford, UK, 1997; Volume 2, pp. 299–321.
59. Pace, C.N. Determination and analysis of urea and guanidine hydrochloride denaturation curves. In *Methods in Enzymology*; Academic Press: Cambridge, MA, USA, 1986; Volume 131, pp. 266–280.
60. Walter, P.; Ron, D. The unfolded protein response: From stress pathway to homeostatic regulation. *Science* **2011**, *334*, 1081–1086. [[CrossRef](#)]
61. Sali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815. [[CrossRef](#)]
62. Guerois, R.; Nielsen, J.E.; Serrano, L. Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. *J. Mol. Biol.* **2002**, *320*, 369–387. [[CrossRef](#)]
63. Whitehouse, D.B.; Putt, W.; Lovegrove, J.U.; Morrison, K.; Hollyoake, M.; Fox, M.F.; Hopkinson, D.A.; Edwards, Y.H. Phosphoglucomutase 1: Complete human and rabbit mRNA sequences and direct mapping of this highly polymorphic marker on human chromosome 1. *Proc. Nat. Acad. Sci. USA* **1992**, *89*, 411–415. [[CrossRef](#)] [[PubMed](#)]
64. Tanguy, A.; Boutet, I.; Boudry, P.; Degremont, L.; Laroche, J.; Moraga, D. Molecular identification and expression of the phosphoglucomutase (PGM) gene from the Pacific oyster *Crassostrea gigas*. *Gene* **2006**, *382*, 20–27. [[CrossRef](#)]
65. Matabos, M.; Plouviez, S.; Hourdez, S.; Desbruyères, D.; Legendre, P.; Warèn, A.; Jollivet, D.; Thiébaud, E. Faunal changes and geographic crypticism indicate the occurrence of a biogeographic transition zone along the Southern East-Pacific Rise. *J. Biogeogr.* **2011**, *38*, 575–594. [[CrossRef](#)]
66. Matabos, M.; Jollivet, D. Revisiting the species' complex of *Lepetodrilus elevatus* (Vetigastropod, Lepetodrilidae) using gastropod samples from the Galápagos and Guaymas hydrothermal vent systems. *J. Moll. Stud.* **2019**, *85*, 154–165. [[CrossRef](#)]

67. Eanes, W.F. Molecular population genetics and selection in the glycolytic pathway. *J. Exp. Biol.* **2011**, *214*, 165–171. [[CrossRef](#)]
68. Verrelli, B.C.; Eanes, W.F. Extensive amino acid polymorphism at the *Pgm* locus is consistent with adaptive protein evolution in *Drosoph. Melanogaster*. *Genet.* **2000**, *156*, 1737–1752. [[CrossRef](#)] [[PubMed](#)]
69. Wheat, C.W.; Watt, W.B.; Pollock, D.D.; Schulte, P.M. From DNA to fitness differences: Sequences and structures of adaptive variants of *Colias* phosphoglucose isomerase (PGI). *Mol. Biol. Evol.* **2005**, *23*, 499–512. [[CrossRef](#)]
70. McDonald, J.H.; Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **1991**, *351*, 652–654. [[CrossRef](#)] [[PubMed](#)]
71. Begun, D.J.; Betancourt, A.J.; Langley, C.H.; Stephan, W. Is the fast/slow allozyme variation at the *Adh* locus of *Drosophila melanogaster* an ancient balanced polymorphism? *Mol. Biol. Evol.* **1999**, *16*, 1816–1819. [[CrossRef](#)]
72. Fontanillas, E.; Galzitskaya, O.V.; Lecompte, O.; Lobanov, M.Y.; Tanguy, A.; Mary, J.; Girguis, P.R.; Hourdez, S.; Jollivet, D. Proteome evolution of deep-sea hydrothermal vent alvinellid polychaetes supports the ancestry of thermophily and subsequent adaptation to the cold for some lineages. *Genome Biol. Evol.* **2017**, *9*, 279–296. [[CrossRef](#)]
73. Bioy, A. Histoire Évolutive et Influence de la Sélection sur la Diversité génétique des Annélides Polychètes D’environnements Extrêmes. Thèse de Doctorat, Sorbonne Université, Paris, France, 2018; pp. 248p.
74. Hoekstra, R.F.; Bijlsma, R.; Dolman, A.J. Polymorphism from environmental heterogeneity: Models are only robust if the heterozygote is close in fitness to the favoured homozygote in each environment. *Genet. Res.* **1985**, *45*, 299–314. [[CrossRef](#)] [[PubMed](#)]
75. Gardner, J.P.A.; Lobkov, I. A test for overdominance at the phosphoglucose-2 locus in Pacific oysters (*Crassostrea gigas*) from B-New Zealand. *Aquaculture* **2005**, *244*, 29–39. [[CrossRef](#)]
76. Levene, H. Genetic equilibrium when more than one ecological niche is available. *Am. Nat.* **1953**, *87*, 331–333. [[CrossRef](#)]
77. Gillespie, J.H. The interaction of genetic drift and mutation with selection in a fluctuating environment. *Theor. Pop. Biol.* **1985**, *27*, 222–237. [[CrossRef](#)]
78. Hedrick, P.W.; Ginevan, M.E.; Ewing, E.P. Genetic polymorphism in heterogeneous environments. *Ann. Rev. Ecol. Syst.* **1976**, *7*, 1–32. [[CrossRef](#)]
79. Hedrick, P.W. Genetic polymorphism in heterogeneous environments: A decade later. *Ann. Rev. Ecol. Syst.* **1986**, *17*, 535–566. [[CrossRef](#)]
80. Watt, W.B. Adaptation at specific loci. I. Natural selection on phosphoglucose isomerase of *Colias* butterflies: Biochemical and population aspects. *Genetics* **1977**, *87*, 177–194. [[CrossRef](#)]
81. Watt, W.B.; Cassin, R.C.; Swan, M.S. Adaptation at specific loci. III. Field behavior and survivorship differences among *Colias* PGI genotypes are predictable from In Vitro biochemistry. *Genetics* **1983**, *103*, 725–739. [[CrossRef](#)] [[PubMed](#)]
82. Carter, P.A.; Watt, W.B. Adaptation at specific loci. V. Metabolically adjacent enzyme loci may have very distinct experiences of selective pressures. *Genetics* **1988**, *119*, 913–924. [[CrossRef](#)]
83. Hartl, D.L.; Dykhuizen, D.E.; Dean, A.M. Limits of adaptation: The evolution of selective neutrality. *Genetics* **1985**, *111*, 655–674. [[CrossRef](#)]
84. Somero, G.N. Temperature adaptation of enzymes: Biological optimization through structure-function compromises. *Ann. Rev. Ecol. Syst.* **1978**, *9*, 1–29. [[CrossRef](#)]
85. Somero, G.N. Proteins and temperature. *Ann. Rev. Physiol.* **1995**, *57*, 43–68. [[CrossRef](#)]
86. Vogt, G.; Woell, S.; Argos, P. Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol. Biol.* **1997**, *269*, 631–643. [[CrossRef](#)]
87. Shin, D.S.; DiDonato, M.; Barondeau, D.P.; Hura, G.L.; Hitomi, C.; Berglund, J.A.; Getzoff, E.D.; Cary, S.C.; Tainer, J.A. Superoxide dismutase from the eukaryotic thermophile *Alvinella pompejana*: Structures, stability, mechanism, and insights into amyotrophic lateral sclerosis. *J. Mol. Biol.* **2009**, *385*, 1534–1555. [[CrossRef](#)] [[PubMed](#)]
88. Bruneaux, M.; Mary, J.; Verhey, M.; Lecompte, O.; Poch, O.; Jollivet, D.; Tanguy, A. Detection and characterisation of mutations responsible for allele-specific protein thermostabilities at the Mn-superoxide dismutase gene in the deep-sea hydrothermal vent polychaete *Alvinella pompejana*. *J. Mol. Evol.* **2013**, *76*, 295–310. [[CrossRef](#)] [[PubMed](#)]