



HAL
open science

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, Anne-Sophie Le Port, Emeline Sabourin, Marie Verheye, Patrice Piccino, Baptiste Faure, Stephane Hourdez, Jean Mary, Didier Jollivet

► To cite this version:

Alexis Bioy, Anne-Sophie Le Port, Emeline Sabourin, Marie Verheye, Patrice Piccino, et al.. 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 (2), pp.206. 10.3390/genes13020206 . hal-03770877v3

HAL Id: hal-03770877

<https://hal.sorbonne-universite.fr/hal-03770877v3>

Submitted on 6 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Balanced polymorphism at the *Pgm-1* locus of the Pompeii worm**
2 ***Alvinella pompejana* and its variant adaptability is only governed**
3 **by two QE mutations at linked sites**

4

5 Bioy Alexis*, Le Port Anne-Sophie*, Sabourin Emeline†, Verheye Marie‡, Piccino Patrice§,
6 Faure Baptiste**, Hourdez Stéphane††, Mary Jean* and Jollivet Didier*

7

8 * UMR 7144 Sorbonne Université-CNRS, Adaptation et Diversité en Milieu Marin, Equipe
9 ABICE, Station Biologique de Roscoff, 29688 Roscoff, France

10 †. Institut de recherche pour la conservation des zones humides méditerranéennes, Tour du
11 Varlat, Le sambuc, 13200 Arles, France. Email: emeline.sabourin@gmail.com

12 ‡ Royal Belgian Institute of Natural Sciences, Rue Vautier 29, 1000 Brussels, Belgium. Email:
13 mverheye@naturalsciences.be

14 § Lycée Lakanal, 3 Avenue du Président Franklin Roosevelt 92330 Sceaux, France. Email:
15 patricepiccino@yahoo.fr

16 **. Biotope – Agence Nord-Littoral, ZA de la Maie, Avenue de l'Europe, 62720 Rinxent,
17 France. Email: bfaure@biotope.fr

18 ††. UMR 8222 CNRS-Sorbonne Université, LECOB, Observatoire Océanologique de Banyuls,
19 66650 Banyuls-sur-Mer, France. Email: hourdez@obs-banyuls.fr

20

21 Corresponding author: Didier Jollivet, jollivet@sb-roscoff.fr, Tel: +33.2.98.29.23.67, Fax:
22 +33.2.98.29.23.24

23

1 Abstract. The polychaete *Alvinella pompejana* lives exclusively on the walls of deep-sea
2 hydrothermal chimneys along the East Pacific Rise (EPR), and displays specific adaptations to
3 withstand the high temperatures and hypoxia associated with this highly variable habitat.
4 Previous studies have revealed the existence of a balanced polymorphism on the enzyme
5 phosphoglucomutase associated with thermal variations, where allozymes 90 and 100 exhibit
6 different optimal activities and thermostabilities. Exploration of the mutational landscape of
7 phosphoglucomutase 1 revealed the maintenance of four highly divergent allelic lineages
8 encoding the three most frequent electromorphs over the geographic range of *A. pompejana*.
9 This polymorphism is only governed by two linked amino acid replacements, located in exon
10 3 (E155Q and E190Q). A two-niche model of selection, including ‘cold’ and ‘hot’ conditions,
11 represents the most likely scenario for the long-term persistence of these isoforms. Using
12 directed mutagenesis and the expression of the three recombinant variants allowed us to test the
13 additive effect of these two mutations on the biochemical properties of this enzyme. Our results
14 are coherent with those previously obtained from native proteins, and reveal a thermodynamic
15 trade-off between protein thermostability and catalysis, which is likely to have maintained these
16 functional phenotypes prior to the geographic separation of populations across the Equator
17 about 1.2 million years ago.

18

19

20 **Keywords:** phosphoglucomutase, balancing selection, thermal stability, gene, adaptive mutations,
21 Alvinellidae

22

1 INTRODUCTION

2
3 A central goal in evolutionary biology is to understand the origin and maintenance of
4 polymorphisms sculpted by natural selection and, more specifically, how the mean pheno- type
5 of a population evolves under heterogeneous and/or changing conditions^[1]. As a consequence,
6 many studies have investigated the maintenance of enzyme polymorphisms by selective
7 processes for species exposed to environmental gradients such as temperature, salinity, or
8 desiccation^[2]. A few decades ago, a series of enzymes interacting in the glycolytic cycle, mostly
9 associated with isomerase and mutase functions, such as phospho-glucose isomerase (PGI),
10 mannose phosphate isomerase (MPI), and phosphoglucomutase, (PGM), were shown to display
11 isoforms that may be the subject of natural selection, leading to habitat-driven differentiation
12 in populations according to temperature, wave action, or metallic pollution^[2–10]. According to
13 Eanes^[11], such branch-point enzymes—which are positioned at the crossroad of metabolic
14 pathways—are likely to be the target of natural selection, as they can orient pathway fluxes
15 according to their protein variation. Among these, alleles encoding the enzyme
16 phosphoglucomutase have been widely studied, with the aim of testing the hypothesis of
17 differential and/or balancing selection. This was mainly achieved by looking at allele^[3,4,12] and
18 heterozygote^[13] frequencies in populations, as well as assessing either the fitness of individuals
19 carrying alleles suspected to be locally advantageous along latitudinal clines^[14] or the kinetic
20 properties of the enzyme isoforms themselves^[6,12].

21 Due to their tremendous thermal variability, caused by the chaotic mixing of cold sea
22 water and hot fluids, hydrothermal vents represent an ideal model for testing the effect of
23 frequent, and unpredictable, spatial and temporal changes of habitats on adaptive enzyme
24 polymorphisms. First, both the fragmentation and instability of the vent discharge likely
25 promote highly dynamic meta-populations with recurrent local extinctions and associated
26 bottlenecks^[15–17]. Long-term oscillations of heat convection beneath the ridge lead to the
27 displacement of the hydrothermal activity along the rift, generating the emergence of new vent
28 sites more or less close to older ones that became extinct, allowing for their rapid
29 recolonization^[16,18]. Such dynamics should have severe implications in reducing the genetic
30 diversity of vent species. Second, variations in temperature, sulphide concentrations, and
31 oxygen concentrations over short periods of time (often ranging from minutes to hours)^[19,20]
32 are likely to affect the respiratory, nutritional, and reproductive physiologies of animals living
33 in such places^[21]. Animals must be able to feed under high temperatures to fuel their symbionts
34 and respire under cold conditions to obtain oxygen. These temporal fluctuations of the vent

1 conditions at the individual scale represents a selective constraint that should promote the
2 maintenance of both thermostable and cold- functioning enzyme alleles through mean
3 overdominance until the exploration of the mutational landscape of a given enzyme leads to the
4 emergence of a highly plastic allelic isoform, and thus enzyme monomorphism. However, the
5 hydrothermal environment is also highly fragmented and heterogeneous, according to the
6 mineral composition of the oceanic crust through which the super-heated fluid moves prior to
7 be expelled above the seafloor^[22]. As a consequence, vent fields often display a mosaic of
8 edifices of different ages^[23,24], whose mean age and size distribution is dictated by the frequency
9 of tectonic and volcanic events, as well as the dynamics of the heat convection beneath oceanic
10 ridges^[25–27]. Vent edifices usually cool down with age until they turn off. Depending on the age
11 of the edifice, populations of vent species are thus spatially subjected to a wide variety of vent
12 conditions, which represents an ecological basis for diversifying selection. As most vent species
13 are quite thermotolerant, such a mosaic of habitats is likely to favour thermostable alleles in
14 newly formed edifices and cold functioning ones in older habitats, the fate of a given allele
15 depending on the relative proportions of these habitats.

16 The polychaete *Alvinella pompejana*, which lives on the hottest part of the hydrothermal-
17 vent environment^[28,29] can withstand temperatures up to 50 °C^[30]. This tube-dwelling worm
18 lives on the walls of hydrothermal-vent chimneys from a latitude of 23° N on the East Pacific
19 Rise (EPR) to 38° S on the Pacific Antarctic Ridge (PAR)^[31]. It has developed peculiar
20 physiological adaptations in order to colonize this hostile habitat^[32,33]. Earlier genetic studies
21 have shown that *A. pompejana* exhibits quite an unusually high level of genetic diversity^[31,34,35]
22 with a non-negligible number of bi-allelic enzyme loci with equally frequent alleles, some of
23 which display different thermal stabilities^[36]. Among them, the enzyme phosphoglucosmutase
24 (PGM-1) possesses four distinct isoforms. Allozymes 90 and 100 have frequencies of
25 approximately 35% and 60%, respectively, in populations of the northern EPR, while the two
26 other isoforms (112 and 78) are rather rare, accounting for the remaining 5%. Although the
27 frequency of allozyme 90 remains constant over the species range, Plouviez et al.^[35] have
28 shown that allozymes 78 and 100 display an abrupt clinal distribution across the Equator, with
29 allozyme 78 becoming the most frequent allele in the southern EPR. Bi-allelism was, thus,
30 preserved all along the EPR despite population isolation, recurrent extinction/recolonizations,
31 and a long history of divergence across the Equatorial barrier.

32 In addition, significant genetic differentiation has been observed between *A. pompejana*
33 populations living in contrasting microhabitats, especially when comparing newly formed ‘still

1 hot' chimneys (i.e., 'hot' niche) to older and colder edifices (i.e., 'cold' niche). The frequency
2 of allele 90 is, indeed, positively correlated with mean temperature at the opening of *Alvinella*
3 tubes and increases in the 'hot' compared to the 'cold' habitat, suggesting that this locus is
4 under diversifying selection^[12]. In vitro experiments on enzyme stability and optima have
5 strengthened this view. They have shown that allele 90 is more thermostable and more active
6 at higher temperatures than allele 100 and, thus, is probably favoured in the 'hot' habitat.

7
8 Although whole-length *Pgm-1* sequences have now been obtained for a large panel of
9 metazoan species, very few studies have been conducted at the population level, most of which
10 involved bacterial strains. While this enzyme has been extensively studied from the 1970s to
11 1990s for adaptive purposes, only few studies have examined the relationship between
12 nonsynonymous changes at the gene level and the subsequent enzymatic performance of
13 alternate isoforms (see, e.g.,^[14,37] for the correspondence between allelism, enzyme thermal
14 resistance, and glycogen storage in *Drosophila*). In this paper, we report a possible case of long-
15 term balancing selection at an enzyme locus where alleles can be maintained by varying
16 selection between two niches, possibly helped by fluctuations in the relative proportions of the
17 two niches over space and time. Most of the documented cases for the long-term persistence of
18 alleles by balancing selection and trans-species poly- morphism come from studies dealing with
19 negative frequency-dependent selection at immune and sex-determination genes^[38,39]. This
20 raises questions about how a chaotic and highly fluctuating two-niche system can promote
21 balancing selection at key branch-point enzymes. The aim of this work was, therefore, to
22 identify the mutations at the basis of the enzyme polymorphism of PGM-1 in the Pompeii worm
23 and to evaluate the effects of these mutations on the thermostability and catalytic efficiency of
24 the enzyme to test whether a trade-off between these two processes is likely to explain the
25 maintenance of the different isoforms. In parallel, we explored the mutational landscape of the
26 gene to search for traces of balancing selection in the vicinity of the amino acid substitutions
27 that led to these isoforms and assess their long-term duration. Finally, we also examined the
28 evolutionary mechanisms: (1) fitness cost to the colonization of newly 'hot' habitats; (2)
29 overdominance associated with the fluctuations of the vent discharge, and (3) the dynamics of
30 hot vs cold habitats (i.e., the two-niches model hypothesis), by which these isoforms could have
31 been maintained in natural populations by analyzing the fecundity of females carrying the most
32 thermostable isoform, and by using MSMS simulations implementing selection across a barrier
33 to gene flow.

34

1 MATERIAL and METHODS

3 Animal sampling

4 Specimens of *Alvinella pompejana* were collected with the ROV Victor 6000 and the
5 deep-sea manned submersible Nautilie during the cruises Phare 2002, Biospeedo 2004 and
6 Mescal 2010 on board of the research vessel L'Atalante. Animals were sampled from targeted
7 sites located on the North EPR (NEPR hereafter) and the South EPR (SEPR hereafter, see Fig.
8 1) over chimneys of different ages ranging from newly formed 'hot' diffusors to large black
9 'smokers', for which thermal and chemical conditions were highly contrasted^[24].

10 In order to test an earlier hypothesis^[12] postulating that individuals carrying genotypes
11 favored during the colonization of newly-formed still 'hot' chimneys may be counter-selected
12 by a lower reproductive fitness under cooler conditions (i.e. trade-off between settlement ability
13 and reproduction), we examined the relationship between PGM-1 genotypes and female's
14 fecundity. To this extent, the size of animals was estimated from the width at the S4 setigerous
15 segment and sexes were determined based on the presence of either a genital pore in females
16 or a pair of sexual tentacles in males^[42]. Mature females collected from both sides of the Equator
17 were dissected to estimate their fecundity per size unit and genotyped at the PGM-1 locus. For
18 each female, the coelomic fluid containing oocytes was carefully removed and resuspended in
19 50 ml of a solution of borate-buffered 3% formalin in seawater. Oocytes were counted
20 following the method previously described by Faure *et al.*^[42] and a one-way ANOVA was
21 performed on size-corrected female fecundities according to the genotype using the software
22 Jamovi (<https://www.jamovi.org>).

24 Identification and characterization of the *AP-Pgm-1* gene

25 *Sequencing of Pgm-1 cDNA using homozygous individuals*

26 Based on allozyme genotypes, 8 homozygous individuals carrying alleles 78, 90 and
27 100 were selected for RNA extractions. Total RNAs were extracted with Tri-Reagent (Sigma)
28 following the manufacturer's instructions and a classical chloroform extraction protocol. Both
29 the quantity and quality of RNAs were assessed with a Nanodrop ND-1000 spectrophotometer
30 (Nanodrop Technologies, Delaware, USA). Five µg of total RNAs were reverse transcribed
31 with a M-MLV reverse transcriptase (Promega), an anchor-oligo(dT) primer (Table S1) and
32 random hexamers (Promega). The reverse anchor and forward nested degenerated PGM
33 primers derived from the oyster *Crassostrea gigas* and human *Pgm-1* sequences were then used

1 to perform the upstream amplification of cDNA fragments (see Table S1). PCR-products
2 containing *Pgm-1* cDNA candidates were then cloned with the TOPO TA Cloning kit
3 (Invitrogen) and, sequenced on an ABI 3130 sequencer using the BigDye v.3.1 (PerkinElmer)
4 terminator chemistry following the manufacturer's protocol. Sequences of clones containing
5 the appropriate *Pgm-1* cDNA fragments were then aligned to reconstruct a series of nearly
6 complete *AP-Pgm-1* cDNA (i.e. only lacking a small part of the 5' end of the coding sequence).

7 8 *Sequencing the Pgm-1 gene with a series of specific exonic primers*

9 Using gDNA, specific reverse primers (Table S1) were also used to amplify the 5'
10 portion of the gene by directional genome walking using PCR^[43]. A series of specific primers
11 were designed based on our cDNA sequences (see Table S1) to amplify both exon and intron-
12 containing portions of the gene with gDNA from the same eight homozygous individuals.
13 Fragments of the gene were obtained using pairs of the least distant forward and reverse primers
14 containing a 6-bp individual identifier (barcode). PCR amplifications were performed in a 25µl
15 PCR reaction volume that comprised 1X buffer (supplied by manufacturer), 2 mM MgCl₂, 0.25
16 mM of each dNTP, 0.4 µM of each primer, 0.5 U of Taq polymerase (Thermoprime plus). The
17 PCR profile included a first denaturation step at 94°C for 4 min followed by 30 cycles at 94°C
18 for 30s, 60°C for 30s and 72°C for 2 min and, a final extension at 72°C for 10 min. All barcoded
19 PCR-products were cloned following the Molecular Cloning Recapture (MCR) method
20 developed by Bierne *et al.*^[44] and sequenced on an ABI 3130 sequencer with the protocol used
21 previously. Alignments of the sequenced fragments allowed us to reconstruct a complete
22 sequence of the *AP-Pgm-1* gene (Accession N° MN218831), its associated cDNA sequence and
23 three native consensus cDNA for the three isoforms (Accession N° MN218832 - MN218839).
24 The analysis of this initial cDNA alignment provided a first information on polymorphic sites
25 between the 3 distinct alleles all along the gene (see Fig. 2).

26 27 *Correspondence between allozymes and non-synonymous mutations of AP-Pgm-1*

28 To examine the correspondence between the only two diagnostic polymorphic non-
29 synonymous EQ mutations found at exon 3 and allozymes 78, 90 and 100, a total of 220
30 individuals were genotyped on the 350bp fragment of the *Pgm-1* exon 3 containing these sites.
31 PGM-1 allozymes were first screened for each individual by electrophoresis on 12% starch-gel
32 at 4°C (100 V, 80 mA, 4 h) with the Tris-citrate pH 8.0 buffer system following the procedure
33 described by Piccino^[12]. The 350 bp exon3 fragment was then amplified by PCR on the same
34 individuals following a gDNA extraction using a CTAB/PVP procedure described by

1 Plouviez^[40]. PCR amplifications were conducted using a specific primer pair (see Table S1)
2 with a first denaturation step at 94°C for 4 min followed by 40 cycles at 94°C for 30s, 60°C for
3 30s and 72°C for 20s and, a final extension at 72°C for 2 min. PCR-products were first double
4 digested on 33 individuals with enzymes *Fai I* (targeting the first substitution site) and *Bsg I*
5 (targeting the second site) as an initial test and then sequenced without cloning on ABI 3130
6 automatic sequencer with the BigDye v.3.1 (PerkinElmer) terminator chemistry after an
7 ExoSAP-IT purification.

8 Forward and reverse sequences were proof-read in CodonCode Aligner to check for the
9 occurrence of single (homozygotes) or double (heterozygotes) peaks at the two polymorphic
10 sites. The allele alignment has been deposited in Genbank (accession N° MN218918-
11 MN219291). Linkage disequilibrium between genotypes, EE, EQ and QE and allozymes 78,
12 90 and 100 was tested using Linkdis^[45] of the software Genetix v.4.05^[46]. The double mutation
13 scoring among individuals allowed us to then estimate heterozygote excesses or deficiencies in
14 populations. Departures to HDW were tested with 1000 permutations of alleles between
15 genotypes using the same software. The exon 3 allele alignment was also used to reconstruct
16 an allelic network using Network 4.5.1.0^[47], in order to examine the permeability of the
17 equatorial barrier between populations at this locus.

18

19 Examining the synonymous and non-synonymous changes along the *AP-Pgm-1* gene

20 Nucleotidic diversities were punctually assessed along the gene by combining direct
21 sequencing and the MCR method on individuals from each side of the EPR (see Fig. 2). These
22 regions included exon 1, exons 4 to 5, end of exon 7 and the beginning of exon 9 (Accession
23 N° MN218840-MN218917 for exon1, Accession N° MN219292-MN219356 for exons 4 and
24 5). In addition, a fragment containing the whole intron 2 and the beginning of exon 3 where the
25 two diagnostic EQ mutations are located (1110 bp) was also sequenced using the MCR
26 method^[44] in order to test whether ‘hot spots’ of mutations occur around these two EQ sites but
27 also to estimate allele divergences (Accession N° MN219357 - MN219404). In the MCR
28 sequence sets, the number of retrieved alleles greatly varied between the different parts of the
29 gene according to the sequencing efficiency and/or cloning success. Artifactual singletons due
30 to the MCR method were removed by comparing the singleton rates between the MCR and
31 direct sequencing datasets on the same fragments.

32 Haplotype diversity (H_d), nucleotide diversity (π) and its synonymous and non-
33 synonymous components (π_S and π_N), and Watterson’s theta (θ_W) were then examined together
34 with deviations to neutral evolution (Tajima’s D and Fu & Li’s F statistics) for both the northern

1 and southern EPR individuals along the gene, using the DNAsp 4.10.3 software^[48] with a
2 sliding window (size = 100 and step = 25). These basic genetic parameters were then compared
3 with the critical values associated with sample size and neutral coalescent simulations, as
4 implemented in the same software. Linkage disequilibrium between segregating sites and
5 recombination among alleles were estimated by calculating the *ZnS* statistics^[49] together with
6 the minimum number of recombination events (*Rm*^[50]). The number of significant associations
7 between linked sites was evaluated following a Fisher's exact test and a Bonferroni correction,
8 implemented in DNAsp 4.10.3. The occurrence of recombinants was also checked using
9 automated RDP and bootscan packages of RDP v.3.44^[51], and recombination hotspots were
10 searched by examining the population recombination rate parameter ($4N_e.r$) along the gene (-
11 recomb and -hotspot outputs) for both the northern and southern populations using Phase 2.1.1
12 software^[52]. Genetic differentiation and allele divergence between the southern and northern
13 parts of the EPR were estimated by calculating *Fst* and *D_{xy}* in DNAsp 4.10.3. Genetic
14 differentiation was tested using 1000 permutations of the sequence data sets using the
15 randomization test developed by Hudson^[53]. Finally, the intron 2-exon 3 alignment (1110 bp)
16 was used to reconstruct a coalescent tree of *AP-Pgm-1* alleles to more specifically evaluate both
17 the intra-locus recombination and allele divergence near the two EQ non-synonymous
18 polymorphic sites. The evolutionary history of alleles was inferred using the Minimum
19 Evolution method implemented in MEGA7^[54] using the NJ algorithm for the initial tree,
20 pairwise deletion of ambiguous sites, and the close-neighbour-interchange (CNI) algorithm.
21 Evolutionary distances were computed using the Maximum Composite Likelihood method.

22

23 Coalescent simulations using models of selection

24 Coalescent simulations under a divergence model with asymmetrical migration rates
25 between two populations ($pop_{1>2}: 2N_e.m = 1$ and $pop_{2>1}: 2N_e.m = 0.1$, $N = 1000$ simulations with
26 $N_e = 50,000$) were performed using MSMS v3.2 software^[55]. Two different hypotheses of
27 balancing selection were evaluated: (1) over-dominance, with genotype selection coefficients
28 $S_{aA} = 500$ and $S_{AA} = 1$ where *S* represents $N_e.s$; and (2) habitat-dependent selection, with four
29 populations and two habitats, where $S_{AA} = 1000$, $S_{aA} = 500$, and $S_{aa} = 0$ in the first habitat and
30 $S_{AA} = 0$, $S_{aA} = 500$, and $S_{aa} = 1000$ in the second habitat. Each set of simulations, including the
31 null hypothesis of asymmetrical migration without selection, were run with two recombination
32 rates ($N_e.r = 1$ or 100). Population parameters, including gene diversities (θ_w , π), and Tajima's
33 *D* within each deme and *Fst* between demes were estimated using the pylibseq 0.2.3 libraries^[56]

1 and a home-made python script.

2

3 Functional and structural analysis of PGM-1 recombinant isoforms

4 *Plasmid construction for enzyme overexpression*

5 Full-length *AP-Pgm* cDNA were obtained from, two homozygous individuals 100/100
6 (EE) and 90/90 (EQ). RT-PCR was conducted with the ClonTech SMARTer Race cDNA
7 amplification kit following the manufacturer instructions and AP-PGMex11 reverse primer (see
8 Table S1). These cDNAs were then used as a target to specifically amplify the complete coding
9 sequence with primers containing cutting sites to be inserted in either Pet20b or PetDuet
10 expression vectors (Table S1). Amplified coding sequences were double-digested with either
11 enzymes *BamHI/NotI* or *AseI/XhoI* in a 25 µl volume containing the restriction buffer, the
12 enzymes, and 1% BSA. The restriction products were then ligated in the appropriate expression
13 vector after purification with a Nucleospin Gel extraction Clean up column (Macherey Nagel)
14 and cloned into BL21DE3 *E. coli* cells amenable for IPTG induction and overexpression.

15

16 *Directed mutagenesis*

17 Using the full-length cDNA with the double mutation EE as a template, mutants ₁₅₅EQ
18 and ₁₉₀EQ were produced by directed mutagenesis following the PCR protocol of Reikofski and
19 Tao^[57]. First amplifications were conducted in 50 µl reaction volume containing: 1X Pfu buffer
20 containing MgCl₂, 0.25 mM of each dNTP, 0.5 µM of each primer (petDuet and mutated
21 primer), 0.5 U of the proof-reading *Pfu* polymerase (Promega) with 30 cycles of 94°C for 30s,
22 60°C for 30s and 72°C for 3 min. Secondly, the two regions of the mutated cDNA were joined
23 following a PCR amplification without primers mixing the two previous PCR-products (1:1)
24 under the same conditions and a final elongation step of 10 min. cDNAs containing the mutated
25 sites ₁₅₅E->Q and ₁₉₀E->Q and the native ₁₅₅E₁₉₀E cDNA were then sequenced on an ABI 3130
26 sequencer with the BigDye v.3.1 (Perkin Elmer) terminator chemistry to verify the sequences
27 before overexpression.

28

29 *Protein expression and purification*

30 *E. coli* (BL21DE3) with the recombinant pETduet plasmid containing either native or
31 mutated PGM cDNA sequences were grown into a LB medium supplied with 100 µg/mL
32 ampicillin at 37°C until they reach an absorbance of 0.6 at 600 nm. Protein expression was
33 induced by adding 1mM IPTG to the medium and kept at 37°C under shaking for 4 hours. Cells
34 were then harvested by centrifugation (4°C/15 000 g/5 min), and the pellets were re-suspended

1 in a binding buffer (20 mM Tris-HCl, pH 6.5, 500 mM NaCl, 5 mM imidazole), disrupted by
2 French Press at 1.6 kbar. After removing cell debris by centrifugation (15 000 g/4°C/60 min),
3 supernatants (1 µg/mL of lysate) were treated with DNase I (Eurogentec) for 1 hour on ice. A
4 first purification step was performed using immobilized metal affinity chromatography with a
5 His-bind resin column (His-Bond kit, Novagen) to recover PGM variants, Protein binding with
6 5 mM and 60 mM imidazole and final elution of allozymes with 1 M imidazole were performed
7 following a classical chromatography protocol (pH 6.5). The eluted fractions were concentrated
8 using 30 kDa molecular cut-off Amicon-Ultra (Millipore™). A second purification step was
9 performed by size-exclusion chromatography (SEC) with Superdex 75 column (1 x 30) (GE
10 Healthcare) at a flow rate of 0.5 mL/min monitored at 280 nm using a 25 mM
11 Na₂HPO₄/NaH₂PO₄, pH 6.5. The purity of proteins was checked by SDS-PAGE stained with
12 Coomassie brilliant blue before being kept at 4°C in an elution buffer supplemented with
13 dithiothreitol (DTT, 10 mM) until use for enzyme assays. The protein concentrations were
14 measured by absorption at 280 nm with the theoretical coefficient of 48,820 M⁻¹.cm⁻¹ as
15 calculated using the ExpASy-ProtParam tool (<http://web.expasy.org/protparam/>).

16

17 *Enzyme activity assay*

18 PGMs activities were assayed by coupling the formation of α-D-glucose 6-phosphate
19 (G6P) from α-D-glucose 1-phosphate (G1P) to NADPH formation using glucose 6-phosphate
20 dehydrogenase (G6PD) as a relay enzyme. The reaction mixture contained 50 mM Tris-HCl,
21 pH 7.4, 0.5 M MgCl₂, 1.2 mM NADP, 0.1 µM G6PD. The recombinant PGMs were used at the
22 following concentrations: [PGM₇₈] = 0.9 µM, [PGM₉₀] = 4 µM and [PGM₁₀₀] = 0.6 µM. The
23 concentration of the substrate (G1P) was varied from 0.2 to 60 mM to determine the kinetic
24 constants K_m and V_{max} using a Lineweaver-Burk plot.

25

26 *Thermal inactivation*

27 The purified PGM activities were measured at 37°C at 340 nm using an UVmc²
28 spectrophotometer (Safas, Monaco) after a 30-minute incubation at challenge temperatures
29 ranging from 5 to 60°C. Activities were then normalized as the percentage of residual activity
30 when compared to the same sample kept in ice. A theoretical curve with the following equation
31 was fitted to each experimental dataset using a nonlinear curve fit algorithm (Kaleidagraph
32 4.5.0, Synergy Software):

$$y = \frac{(y_N + m_N \cdot T) + (y_D + m_D \cdot T) \cdot \exp\left(\frac{m(T - T_m)}{RT}\right)}{1 + \exp\left(\frac{m(T - T_m)}{RT}\right)} \quad [58] \quad (1)$$

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

7

8 *Guanidinium chloride-induced unfolding of PGM isoforms*

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

$$19 \quad f(I) = (y_N - y) / (y_N - y_I)$$

$$20 \quad f(II) = (y_I - y) / (y_I - y_U)$$

21 where y_N , y_I and y_U are the measured fluorescence intensity respectively of the native,
22 intermediate and unfolded state and y the fluorescence intensity observed at a given GdmCl
23 concentration. The unfolded fractions $f(I$ or $II)$ data were plotted against GdmCl concentrations
24 and theoretical curves, defined by the following equation, have been fitted on the experimental
25 dataset using a nonlinear curve fit algorithm (Kaleidagraph 4.5.0, Synergy Software),

$$26 \quad f(I \text{ or } II) = \frac{\exp\left(-m \frac{(C_m - [GdmCl])}{RT}\right)}{1 + \exp\left(-m \frac{(C_m - [GdmCl])}{RT}\right)} \quad [60] \quad (2)$$

27 where T is the absolute temperature, R is the universal gas constant, C_m is the concentration of
28 GdmCl at the midpoint of the transition, m the dependence of the Gibbs free energy of unfolding
29 reaction (ΔG) on the denaturation concentration of GdmCl. Knowing C_m and m , standard Gibbs
30 free energy of the unfolding reaction in absence of denaturant, ΔG_{H2O}° , can be calculated
31 according to the relation:

1
2 $\Delta G^0_{H2O} = m C_m$ [58] (3).
3

4 *3D PGM Structure Modelling*

5 PGM 78, 90 and 100 3D protein conformations were modelled with the Modeller
6 9v13^[61], using the structure of the crystallized rabbit phosphoglucomutase with its substrate α -
7 D-glucose 1-phosphate as a template (pdb file 1C47). This protein comprises 561 amino acids
8 with a resolution of 2.70 Å that shares 65% sequence identity with that of *Alvinella pompejana*.
9 One hundred models were generated for each PGM isoform and their quality was assessed using
10 the Modeller Objective Function parameter. Finally, a structure optimization was obtained
11 using the repair function of the FoldX software^[62].
12

13 RESULTS

14 Sequencing *AP-Pgm-1* cDNA from homozygous genotypes

15

16 Full-length *Pgm-1* cDNA sequences were obtained from three genotypes 100/100, three
17 genotypes 90/90 and only two genotypes 78/78. This led to a complete cDNA sequence of 562
18 codons without indel between alleles encoding the three distinct allozymes (Fig. S1). The
19 consensus protein sequence fell into the phosphoglucomutase 1 family of proteins with a blastp
20 e-value of 0.0 (65-72% of identity over 99% of 562 residues with the sequence from the oyster
21 *Crassostrea gigas*, and a selection of vertebrate species). Out of the 16 cloned sequences, only
22 two non-synonymous mutations on exon 3 allowed us to discriminate the three main genotypes
23 (100/100, 90/90 and 78/78). These polymorphic mutations corresponded to the replacement of
24 a glutamic acid (E) by a glutamine (Q) at positions 155 and 190. Another replacement of a
25 valine (V) by a leucine (L) at position 40 was also found in exon 1 at intermediate frequency,
26 but this amino-acid polymorphism was not linked to a given electromorph. A phenylalanine (F)
27 replacement by a leucine (L) was also found at position 502 in cDNA encoding allozyme 90.
28

29 Assignment of the two EQ amino-acid replacements to allozymes in natural populations

30

31 In order to address the relationship between the two QE substitutions depicted from the
32 cDNA sequences and allozymes, direct sequencing (and/or RFLP) were performed on a portion
33 of exon 3 (94 codons) containing the double diagnostic mutations EQ in 220 individuals from
34 both sides of the East Pacific Rise previously genotyped at the PGM-1 enzyme. The linkage

1 disequilibrium between the two EQ mutations at codon positions 155 and 190, and allozymes
2 was highly significant (Table 1) with correlation coefficients (R_{ij}) greater than 70% (p-
3 values < 0.0001). This provides a very reliable correlation in which combinations QE, EQ and
4 EE correspond to the isoforms 78, 90 and 100, respectively. The most negatively-charged
5 allozyme 112, which is rare and always found at the heterozygous state in the northern
6 populations was also assigned to genotype EE, suggesting that an additional replacement is
7 occurring elsewhere in the protein. From this genotyping, groups of individuals from either the
8 North or the South EPR did not depart significantly from the Hardy-Weinberg proportions.
9 However, observed and expected heterozygosities were both greater in the northern population
10 ($H_{o-North}$: 0.40 vs $H_{o-South}$: 0.29). Interestingly, allele QQ was not found in any of the
11 populations. The frequencies of EE, EQ and QE alleles at the sampled localities are summarized
12 in Table S2. A more thorough analysis of the North/South genetic differentiation was conducted
13 on the 374 allelic sequences obtained by direct sequencing (see alignment in supplementary
14 data). The resulting haplotype network (Fig. S2) shows a quasi-complete isolation of the
15 northern and southern populations with a F_{st} value of 0.510 (see Table 2). Based on the 282 bp
16 alignment, PGM90 (EQ) found in the Southern population derives directly from the northern
17 PGM90 (EQ) by one fixed mutation and the southern PGM78 (QE) differs by 2 mutations from
18 the northern PGM100 (EE). The haplotype network also indicated that at least three alleles
19 sampled in the southern populations originated from the northern populations, suggesting that
20 the barrier to gene flow is not completely sealed.

21

22 Cryptic amino-acid variation along the *AP-Pgm-1* gene

23

24 The full sequence of the *AP-Pgm-1* gene with the location of polymorphic codons and
25 primers are shown in Fig. S1. The total length of the nucleotidic sequence is 4372 bp. The gene
26 is subdivided into 9 exons and 8 introns which length ranges from 155 to 848 bp. The coding
27 sequence of 1686 bp (562 codons) has an overall GC content of 43.5% (compared to only 29.3%
28 in the intronic regions). When compared to human and oyster *Pgm-1* genes^[63,64], the largest
29 *AP-Pgm-1* exon, comprising 156 codons (other exons vary from 40 to 81 codons), corresponds
30 to the fusion of exons 3, 4, and 5 of the human *Pgm-1*. This fusion is shared with the oyster *C.*
31 *gigas*, suggesting that annelids and mollusks are sharing the same gene architecture (Fig. 2).

32 Besides the two QE changes affecting the net charge of the protein in exon 3, other, less
33 common, cryptic amino-acid replacements were found along several regions of the *AP-Pgm-1*
34 CDS. This allowed us to estimate gene diversities and the south/north divergence over an

1 overall portion of about 3 kb (two thirds of the gene, see Table 2). Gene diversities were almost
2 constant over the *AP-Pgm-1* gene but allele divergence increases locally in the vicinity of the
3 two segregating EQ sites (Table 2). Looking more closely at the site variation along the gene
4 using a sliding window on our set of sequenced fragments indicated that gene diversity is also
5 slightly higher in exon3 where the two QE substitutions are found with values almost identical
6 to those depicted in intron2 (Fig. 3). This slight increase corresponded to peaks of positive
7 Tajima's D values, which raised up to +0.5 at the beginning of exon3, suggesting that the
8 presence of the two linked non-synonymous mutations may be associated with a hotspot of
9 gene diversity. Observed genetic diversities as estimated from θ_w and π were however not
10 significantly greater than expected from neutral coalescent simulations for both the southern
11 and northern populations over all the investigated *Pgm1* fragments (Fig.3, Table 2). Together
12 with the two QE variant sites, the genotyping of exon 1 also confirmed the occurrence of a
13 trans-equatorial V40L substitution found at a frequency of 0.15 restricted to the southern EQ
14 allele (PGM90) and one of the two northern allelic lineages, irrespective of the mutations EE
15 (PGM100) and EQ (PGM90). The direct sequencing of the two other genic regions located
16 either between exons 4 and 5 and between exons 6 and 8 did not show any additional diagnostic
17 amino-acid changes between the 3 allelic lineages EE, EQ and QE. By contrast, several
18 synonymous changes and indels appear to segregate between different allelic lineages along the
19 gene (see sequence alignments provided as supplementary data for exons 1, 3, 4, 5, 7 and introns
20 2, 6, and 7).

21

22 Estimating allele divergence and linkage disequilibria between segregating sites

23

24 To examine more specifically allele divergence and linkage disequilibria between
25 segregating sites within allelic lineages, a sequencing of recaptured alleles was targeted on the
26 longest region of the *AP-Pgm-1* gene (1110 bp). This region containing intron 2 and the two
27 allozyme-diagnostic sites EQ on exon 3 was thus genotyped from a subset of individuals. The
28 sequencing of 48 alleles highlighted the presence of a high level of synonymous polymorphism
29 with a strong linkage disequilibrium between these sites (Table 2), and two diagnostic indels in
30 intron 2 (insertions referred to as A and B following their order in the intron). These segregating
31 sites and indels allowed us to determine 4 divergent allelic lineages with a few recombinants
32 between them. These alleles were split between the northern and southern populations. In the
33 southern population, the two allelic lineages L1 and L2 refer to the allozyme-diagnostic double
34 mutation QE and EQ, whereas allelic lineages L3 and L4 refer to EQ and a mixture of EQ and

1 EE in the northern population (Fig. 4). At least, 9 and 11 synonymous mutations were fixed in
2 intron 2 between allelic lineages L1 and L2, on one hand, and L3 and L4, in the other hand.

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

7 In the Northern population, the two divergent lineages L3 and L4 also display linked
8 sites with either the A indel (L1) or the B indel (L2) but these two lineages are not completely
9 associated with the double mutations EE and EQ. Alleles EE were only found in one of the two
10 lineages and one recombinant between L3 and L4, suggesting that these two lineages have
11 recombined once (Fig 4; Table 2). Alternatively, allele EE could derive from one of the two
12 lineages.

13 To estimate the recombination rate, we examined the distribution of the *Rho* parameter
14 4N.r with Phase 2.1.1 over a greater proportion of the gene (1-2860 bp) using segregating sites
15 (n=53) shared between northern and southern individuals that were successfully sequenced for
16 all exon-intron fragments of the *AP-Pgm-1* gene (N=20). Results from the Phase -recomb and
17 -hotspot outputs clearly indicated that the recombination rate between alleles remains extremely
18 low all along the gene (average local *Rho*= 0.033 and 0.008 for the northern and southern
19 populations, respectively, which further increased to nearly 2 in the southern population at the
20 end of the gene near the position 2140. This study therefore indicated that the 4 allelic lineages
21 greatly diverge one to each other in the vicinity of the double mutation characterizing
22 allozymes, with divergence even greater between allelic sequences of the same population (0.7-
23 1%) than those of the two geographic regions investigated (0.9%).

24
25 To test whether the *AP-Pgm1* genetic patterns may be maintained by selection,
26 population parameters of both the northern and southern populations were simulated using a
27 msms structured coalescent with and without selection. Simulations indicated that a low
28 asymmetrical migration across the equatorial barrier with low or no recombination does not
29 explain by itself the observed patterns of genetic diversities found for the *AP-Pgm-1* gene
30 (Table S3). Simulated *Fst* values were around 0.8 and asymmetrical deme diversities were at
31 least two times smaller than the observed ones ($\pi = \theta_w = 3$) with and without recombination. In
32 this context, Tajima's *D* was close to zero within each deme as observed but highly positive
33 (+2.75) for the overall population when the observed one was also close to zero. Introducing
34 selection led to a better fit of simulated parameters to the observed ones. Simulations with

1 overdominance and low recombination led to a slight decrease of F_{st} values (0.7) between
2 demes, an increase of the within-deme genetic diversities close to the observed ones but also
3 produced greater positive Tajima's D (+1.3 for each deme). Our best fit to values observed in
4 the worm's populations was obtained with the two-niches model simulations ($F_{st}=0.45$,
5 converging nucleotidic diversities ($\theta_w=17 \rightarrow 13$) and Tajima's D (+0.8 \rightarrow +0.4) estimates within
6 and between demes). These simulated values were even closest to those estimated in the vicinity
7 of the two EQ sites (intron2, see Tables 2, S3).

9 Conformational stability, thermal inactivation and kinetics of the mutated isoforms

10
11 The obtention of full-length *AP-Pgm-1* cDNAs allowed us to examine the direct effect
12 of the two QE substitutions on the thermal stability and efficiency of the PGM-1 enzyme using
13 *in vitro* directed mutagenesis. To determine the conformational stability of the three
14 recombinant isoforms of the PGM-1, their guanidinium chloride (GdmCl)-induced unfolding
15 was studied. Variations of fluorescence with increasing concentrations of GdmCl were biphasic
16 (Fig S3) suggesting that the protein follows a three-state model of denaturation. For each
17 transition, the unfolded fraction of protein (f_u) was determined (Fig S4) and Gibbs free energy
18 change associated with each transition calculated (Table 3). For the two transitions, the PGM90
19 (EQ) appears more stable than the two other isoforms. PGM78 (QE) appears more stable than
20 PGM100 for the first transition ($\Delta G^{\circ}_{H_2O} = 8.0 \pm 0.46 \text{ kJ.mol}^{-1}$ vs. $\Delta G^{\circ}_{H_2O} = 6.06 \pm 0.81 \text{ kJ.mol}^{-1}$
21 respectively), but not for the second transition ($\Delta G^{\circ}_{H_2O} = 15.43 \pm 0.93 \text{ kJ.mol}^{-1}$ vs. $\Delta G^{\circ}_{H_2O} =$
22 $15.13 \pm 0.98 \text{ kJ.mol}^{-1}$ respectively). The T_m values obtained from the theoretical curve fitted on
23 the thermal inactivation experimental data (inset Fig. 5) are very similar for PGM78
24 ($46.5 \pm 1.7^{\circ}\text{C}$) and PGM100 ($44.0 \pm 0.1^{\circ}\text{C}$), but markedly higher for PGM90 ($50.9 \pm 0.7^{\circ}\text{C}$).

25 Enzyme kinetic analyses of the three PGM isoforms are also presented in Table 3. The
26 catalytic efficiency of the PGM78, evaluated by the ratio k_{cat}/K_m^{app} , appeared 125- and 70-fold
27 higher than that of PGM90 and PGM100, respectively. Both changes in K_m^{app} (for the substrate
28 Glucose 1 phosphate (G1P)) and k_{cat} , explain most of the difference in the catalytic efficiency
29 of the three isoforms. K_m^{app} (G1P) and k_{cat} of the PGM78 are indeed respectively tenfold lower
30 and a tenfold higher than that of the two other isoforms (see Table 3).

31 Fitness cost of individuals carrying the thermostable allele in terms of female fecundity

1 The Pompeii worm females exhibited an average coelomic fecundity of 200,000 oocytes
2 with a great variability among them (values ranged from 1200 to 450 000 oocytes depending
3 on size (age) and the reproductive state^[42]). As opposed to our expectations, females carrying
4 the allele 90 were on average more fecund than homozygous females carrying alleles 78 and
5 100. However, distributions of fecundity corrected by the size of the female were not
6 significantly different one to each other according to *Pgm-1* genotypes (One-way ANOVA:
7 $F=1.08$, $p=0.37$), see Fig. S5). This finding clearly indicates that the ability to live under hotter
8 conditions is not counter-balanced by a lesser reproductive success, at least for the females.

10 DISCUSSION

11
12 Based on allozyme data, Piccino et al.^[12] previously proposed that the enzyme
13 polymorphism of the Pompeii worm *A. pompejana* may be balanced at the locus *Pgm-1*, at least
14 in populations of the northern EPR. They showed that allozymes 90 and 100, indeed, display
15 distinct thermal stabilities and kinetic optima, with the frequency of the most thermostable
16 isoform (allozyme 90) being positively correlated with temperature in newly formed
17 edifices^[12,36]. As the Pompeii worm is the only vent species capable of colonizing newly
18 formed, still-hot hydrothermal chimneys, bearing thermostable alleles is likely to represent an
19 adaptive advantage. The maintenance of polymorphism by selection on thermostable alleles,
20 however, remains unresolved. If advantageous in the hottest part of the vent environment,
21 thermostable alleles can be expected to spread rapidly in the population through recurrent
22 selective sweeps. This is obviously not the case for the *Pgm-1* locus, which exhibited three
23 major isoforms of different thermal stabilities (allozymes 78, 90, and 100), showing sharp
24 frequency differences across the equatorial barrier to gene flow, as depicted by Plouviez et
25 al.^[40]. Several hypotheses have thus been proposed by Piccino et al.^[12], regarding the
26 maintenance of alleles at the PGM-1 enzyme. These includes:

- 27 (i) allele over-dominance due to the rapid alternation of aerobic/anaerobic vent conditions;
- 28 (ii) a fitness cost for individuals carrying the most thermostable allele at this locus; and
- 29 (iii) a two-niche model effect, due to fluctuating proportions of ‘hot’ and ‘cold’ habitats along
30 the EPR.

31 In the present study, we sequenced the three major *Pgm-1* alleles of the worm to
32 investigate the distribution of non-synonymous polymorphisms along the gene, examine their
33 relationship with allozymes, and assess their evolutionary fate. We demonstrated that only two
34 linked mutations ($E_{155}Q$ and $E_{190}Q$) are associated with the net charge of allozymes and are also

1 responsible for the thermal performance of the three allozymes (78, 90, and 100). Examining
2 the evolutionary history of these alleles indicated their rather old origin, which predates the
3 vicariant event that separated the EPR vent fauna across the Equator about 1.2 million years
4 ago^[40,65,66]. In the following discussion, we therefore examine arguments towards the adaptive
5 maintenance of PGM-1 isoforms. We propose that thermal compensation represents a powerful
6 mechanism by which different enzymatic properties might be maintained under balancing
7 selection, at least, during the exploration of the mutational landscape of the protein that will
8 lead to the emergence of the ‘optimal’ isoform, in order to optimize metabolic fluxes, as
9 previously stated by Eanes^[67].

11 *Two-allele polymorphism at the Pgm locus: a long story of balancing selection?*

12 The non-synonymous polymorphisms associated with the four allelic lineages of *AP-*
13 *Pgm-1* appeared to be quite low ($\pi_N = 0.0025$, on average). Such a result sharply contrasts with
14 earlier studies on branch-point glycolytic enzymes (that control the metabolic flux for transport,
15 storage, and breakdown of carbohydrates), for which numerous cryptic non-synonymous
16 changes were described^[67]. High levels of gene diversity have indeed been observed between
17 the slow, medium, and fast electrophoretic *Pgm-1* alleles of *Drosophila melanogaster*^[14,68], and
18 between phosphoglucose isomerase (PGI) alleles of *Colias* butterflies^[69], suspected to evolve
19 under balancing selection. By contrast, only eight nonsynonymous mutations (E₃₇Q, V₄₀L,
20 E₁₅₅Q, E₁₉₀Q, R₃₄₃I, G₃₅₈S, T₃₆₆M, and F₅₀₂L), some of which were at low frequency, were
21 detected by the direct comparison of allelic sequences of *A. pompejana*. Moreover, even if gene
22 diversity was slightly higher in the intronic region preceding exon 3 and the exon 3 itself where
23 the two EQ sites are found, its variation along the gene does not fit perfectly with the
24 expectations of long-term balancing selection. A weak hotspot signal of silent site variation,
25 supported by slightly positive Tajima’s D and Fu & Li’s F statistics was, however, observed
26 near the doubly selected sites E¹⁵⁵Q/E¹⁹⁰Q, but not as strong as signals depicted for the *Adh*
27 locus in *Drosophila*^[70,71]. Theoretical effects of balancing selection on nearby genome regions
28 promote increase in genetic diversity near the selected site due to lack of recombination and
29 the long-term accumulation of mutations^[38]. The very low level of nucleotidic polymorphism
30 found at *AP-Pgm-1* can, however, be partially explained by recurrent population bottlenecks,
31 due to the challenging environmental conditions that affect the whole vent fauna^[15]. The joint
32 action of abrupt demographic changes and habitat specialization should, indeed, promote
33 enzyme monomorphism. Under such conditions, the level of polymorphism observed at *Pgm-*

1 *I*, although low, appears to be quite unusual when compared to most of the genes examined in
2 *A. pompejana*. In alvinellid worms - and, especially, thermophilic species - proteins are under
3 strong purifying selection, with transcriptome-wide d_N/d_S averages very close to zero, and
4 individual gene values ranging between 0.02 and 0.05^[72]. To this extent, it is worth noting that
5 gene diversity at the *Pgm-1* locus appears to be locally two to four-fold higher than the genome-
6 wide average value (ddRAD overall $\pi = 0.0025$ ^[73]), and from other reported genes^[35].

7 Looking more specifically at the four allelic lineages of *AP-Pgm1* near the EQ sites
8 (intron 2) clearly indicates that they have accumulated a great number of synonymous
9 substitutions due to their separation with almost no recombination events (low values of R_m
10 and Rho , see the Phase 2.1.1 analysis). The two allelic lineages, L1 and L2, present in the
11 southern population exhibited 1% divergence between them, with a strong linkage
12 disequilibrium between the two variant sites E¹⁵⁵Q¹⁹⁰(PGM90) and Q¹⁵⁵E¹⁹⁰(PGM78), and the
13 silent substitutions found in intron 2. This suggests that the two allelic lineages evolved
14 separately, without recombination in the vicinity of the two nonsynonymous sites, for a long
15 period of time. The two northern allelic lineages (L3 and L4) also diverged by 0.7% and display
16 two diagnostic indels in intron 2. These silent mutations and indels are also linked, suggesting
17 once more that these two lineages evolved separately, but for a shorter period of time. However,
18 diagnostic mutations were not completely linked to the variant sites E¹⁵⁵E¹⁹⁰(PGM100) and
19 E¹⁵⁵Q¹⁹⁰(PGM90). Although L3 forms a single clade associated with E¹⁵⁵Q¹⁹⁰, L4 is a mixture
20 of E¹⁵⁵E¹⁹⁰ and E¹⁵⁵Q¹⁹⁰, suggesting either that at least one recombination event occurred
21 between the two northern alleles, or that E¹⁵⁵E¹⁹⁰(PGM100) is a recently derived variant in the
22 northern population. Finally, both divergences observed between alleles within each population
23 were of the same amplitude as the divergence estimated between the southern and northern
24 alleles (0.9%), which coincided with an overall significant F_{st} value of 0.588 between them. If
25 we accept that the southern (L1 and L2) and northern (L3 and L4) allelic lineages became
26 isolated after the appearance of the physical barrier to dispersal - about 1.2 million years
27 ago^[35,40] - this clearly indicates that the co-occurrence of the four highly divergent *Pgm-1* alleles
28 derives from an older polymorphism, predating the vicariant event that separated the northern
29 and southern vent fauna of the East Pacific Rise, with the possible emergence of the genotype
30 E¹⁵⁵E¹⁹⁰ in the north. Such a scenario is likely confirmed by the distribution of alleles in the
31 exon 3 haplotype network and the signature of the linked silent polymorphic sites in introns 1

1 and 2, where northern alleles seem to derive from a southern allele bearing the Q^{155E}¹⁹⁰
2 mutation.

3

4 *Selective modalities for the maintenance of a balanced polymorphism*

5 The long-term evolution of *Pgm-1* alleles without recombination - at least, in the first
6 part of the gene - and their frequency changes according to environmental conditions, raises
7 questions regarding the selective modalities acting on the co-occurrence of alleles when one of
8 the two alleles is better adapted to high temperatures^[12]. One of the first explanations for the
9 adaptive maintenance of AP-PGM1 isoforms was to consider that the rapid alternation of oxic
10 and anoxic conditions during venting should favour heterozygote excesses if the heterozygote's
11 fitness is close to that of the favoured homozygote in one of the two habitats^[74]. Pogson^[13]
12 previously proposed that over-dominance represents the most likely evolutionary mechanism
13 at the origin of the maintenance of a balanced polymorphism at the *Pgm-2* locus for the oyster
14 *C. gigas* (see also^[75], for its link with the individual's growth rate). Here, we were not able to
15 detect any over-dominance at the *Pgm-1* locus in terms of heterozygote excesses in natural
16 populations. Simulations of structured coalescent with asymmetrical migration and over-
17 dominance always led to very high within-deme positive Tajima's D and unequal diversities
18 between overall and within demes (not found in the observed dataset). This finding confirms a
19 previous study^[12] considering allozymes. However, it is worth noting that such an advantage
20 can be easily masked by the temporal dynamics of the thermal habitat (i.e., chimneys refreshing
21 with time) and the juxtaposition of chimneys of different ages. The worm is, indeed, exposed
22 to a mosaic of fluctuating thermal habitats where temperature could vary spatially, according
23 to the age of the chimneys^[12].

24 Maintenance of allozymes with different thermal stabilities can be also explained by a
25 two-niche model of local differentiation between habitats and drift^[76-79]. Simulating
26 coalescences with a two-niche model and a theta value equal to the observed value provided
27 parameter estimates (Fst, as well as diversities and Tajima's D) much closer to values observed
28 in the vicinity of the two EQ sites than the two other models of asymmetric gene flow with and
29 without over-dominance. Given the spatial and temporal dynamics of the hydrothermal
30 discharge, changes in the frequency of *Pgm-1* alleles could be either due to local selection or
31 exacerbated genetic drift associated with the dynamics of colonization of the newly opened
32 sites. Indeed, the dynamic nature of hydrothermal vents over longer time scales (i.e., years) led
33 to a very patchy and transient habitat scattered along the EPR, with a complex heterogeneity of

1 age-driven vent conditions. This can be seen as a multitude of distinct ecological niches for the
2 same species. In this context, the proportion of newly formed ‘still hot’ chimneys and older
3 ‘colder’ ones greatly vary over time. These dynamics depends on the spreading rate of the rift
4 and, thus, the frequency of tectonic and volcanic events along the East Pacific Rise.
5 To explain the maintenance of a bi-allelic polymorphism at the *Pgm-1* locus, Piccino et al.^[12]
6 proposed a fitness cost of worm colonization during the early stages of a chimney. The first
7 settlers on ‘hot’ (>100 °C) anhydrite chimneys benefit from a lack of predators and competitors.
8 Colonists may, therefore, display more thermoresistant alleles but lower reproductive
9 investment and/or survival. Watt et al.^[80–82] and Wheat et al.^[69] previously showed that both
10 PGI and PGM have a great contribution to the mating fitness of male *Colias* butterflies, likely
11 as the result of longer and more vigorous flight within the day. However, based on female
12 fecundity, we were not able to observe fitness differences between the *Pgm-1* genotypes. This
13 suggests that a better predisposition to colonize still ‘hot’ chimneys is probably not
14 compensated by reduced reproductive success to prevent fixation of the advantageous allele. In
15 fruit flies, the *Pgm* locus represents a quantitative trait for glycogen storage and, hence, the
16 ability to survive better under starvation^[34]. In the case of *A. pompejana*, differences in the
17 thermal regime could be great between colonists and reproducers. As a consequence, colonists
18 subjected to longer periods of high temperature (and associated hypoxia) may be maladapted
19 to produce and use their glycogen reserves. This should have consequences in their investment
20 into reproduction. To the contrary, secondary settlers arriving in much cooler conditions are
21 more likely to use their glycogen reserves to massively invest in the production of gametes, as
22 was previously shown by Faure et al.^[42]. Colder conditions seem to be a prerequisite for
23 releasing fertilized eggs after pairing, as embryos are not able to develop at temperatures greater
24 than 15 °C^[41].

25

26 *Adaptive polymorphism: a trade-off between enzyme thermostability and catalysis*

27 In *Drosophila*, *Pgm-1* variants play a non-negligible role in regulating the metabolic
28 energy pool along latitudinal clines, where a decrease in temperature is compensated for by an
29 increase in enzyme activity. Populations of *D. melanogaster* living at the highest (and, thus,
30 coldest) latitudes possess PGM allozymes with a higher catalytic efficiency and greater
31 glycogen contents. According to the theory of metabolic flux, this can be an adaptive means of
32 temperature compensation to maintain the same glycogen contents over the latitudinal
33 gradient^[67]. Differences in both protein thermostability and catalytic efficiency between *Pgm*

1 alleles have been previously reported to explain both local differentiation and latitudinal clines
2 in the oyster *C. gigas*^[6,13] and *D. melanogaster*^[34]. By comparison, thermal compensation may
3 be directly linked to different biochemical phenotypes that interact with the growth rate and
4 reproductive effort of the worms. The theory predicts that differences in activity at only one
5 enzyme must be substantial to affect metabolic fluxes between genotypes^[83]. To test this
6 hypothesis, one could measure the effect of non-synonymous mutations on the functional
7 properties of the enzyme, its conformational stability, and their effects on population fitness. In
8 this study, three recombinant isoforms of PGM-1 - E¹⁵⁵E¹⁹⁰(PGM100), E¹⁵⁵Q¹⁹⁰(PGM90), and
9 Q¹⁵⁵E¹⁹⁰(PGM78) - were obtained by directed mutagenesis. The replacement of the glutamate
10 by a glutamine at position 190 increased the conformational stability and thermostability of the
11 protein, confirming that PGM90 is the most thermostable isoform. As discussed by Piccino et
12 al.^[12], carrying this allele may be advantageous during the colonization of newly formed
13 chimneys, whose surface temperature usually exceeds 50 °C. Unexpectedly, PGM90 exhibited
14 a decrease in the catalytic efficiency of the enzyme when compared to the two other recombined
15 variants (*k_{cat}/K_m* was a hundred times for PGM78 and nearly two-fold greater for PGM100,
16 when compared to PGM90). The recombinant PGM90 also exhibited the lowest affinity for its
17 substrate, glucose-1-phosphate, at 17 °C. This finding is of importance, as such a genetically
18 determined trade-off between protein stability and enzyme activity has not, so far, been reported
19 in other invertebrate species subjected to balancing selection^[6,13,34,82]. Increased thermostability
20 of a protein is often associated with a decrease in the flexibility of the molecule and, thus, the
21 dynamics of the enzyme reaction^[84,85]. Our results are in perfect agreement with these
22 theoretical expectations. They support the positive role of a thermodynamic trade-off between
23 thermostability and catalysis, as previously proposed by Eanes^[11], to explain the co-occurrence
24 of alleles. PGM90 can remain stable for a longer period of time but is less efficient in either
25 producing or consuming the glycogen reserves of the worm than the two other isoforms. To
26 this extent, the fact that the *k_{cat}/K_m* ratio of isoform PGM78 is much higher than that of the
27 isoform PGM100 can explain why isoform 78 is more frequent in the southern populations
28 (about 80%), when compared with isoform PGM100 in the northern populations (around 70%).
29 The balance between allele frequencies from both sides of the Equator may be dictated by the
30 selective coefficient attributed to each genotype as a direct reflection of the catalytic efficiency
31 difference between PGM90 and its alternative isoforms.

32

33 *Structural effect of mutations E155Q and E190Q*

1 The location of the two main polymorphic sites (E¹⁵⁵Q and E¹⁹⁰Q) in the 3D model
2 structure of phosphoglucomutase 1 expose them to solvents, and are not located in the binding
3 domains of the enzyme (Figure 6). Their potential effect on the catalytic properties of the
4 enzyme is, therefore, not the result of a direct interaction with the substrate and/or the residues
5 involved in catalysis. This is not surprising, as most of the mutations affecting the binding of
6 the substrate glucose-1-phosphate, the Mg²⁺ ion, and phosphate should be deleterious.
7 Similarly, in a study of the polymorphism of the enzyme PGM in *D. melanogaster*, none of the
8 21 polymorphic amino acid replacements were located in the catalytic site of the enzyme [34].
9 Based on their location, both substitutions should affect the net charge of the protein in the
10 same way. However, the isoforms 78 and 90 did not have the same electrophoretic mobility,
11 suggesting that some post-translational modification may be involved in the electrophoretic
12 separation of the three isoforms. The gain of a glutamate at position 155 may be associated with
13 a potential ionic bond with histidine 157, with a distance of 6.5 Å between them. Ionic and
14 hydrogen bonds have been shown to increase the stability of enzymes^[86] and partially explain
15 the thermostable 3D structure of the Cu–Zn and Mn superoxide dismutase enzymes in *A.*
16 *pompejana*^[87,88]. This may account for the increased thermostability of isoform 90 but should
17 also have the same effect for isoform 100, which was obviously not the case. This suggests a
18 negative effect of glutamate (E) at position 190, which negates the positive effect of glutamate
19 at position 155. Alternatively, the 3D model comparison of the three isoforms shows that the
20 replacement of one glutamine (Q) by a glutamate (E) at position 190 (as in allozymes 78 and
21 100) introduces a negative charge in a region already enriched in acidic residues. This high
22 density of negative charges could have a destabilizing effect on the protein structure through a
23 Coulomb repulsion effect, thus potentially leading to greater sensitivity to temperature. Finally,
24 the glutamine replacement at position 155 is also likely to play a key role in the molecular
25 dynamics of the protein, especially during the 180° rotation of the reaction intermediate
26 (glucose-1,6-diphosphate) inside the active site. This potentially explains the higher enzymatic
27 efficiency of the isoform 78.

28

29 **CONCLUSION**

30

31 Exploration of the mutational landscape of phosphoglucomutase 1 revealed the maintenance of
32 four divergent allelic lineages over the geographic range of *A. pompejana*. The enzyme
33 polymorphism, only governed by two linked amino-acid replacements (E¹⁵⁵Q and E¹⁹⁰Q), is

1 likely to be maintained by balancing selection, and better fits with a two-niche model of
2 selection, in which ‘cold’ and ‘hot’ conditions alternate. These isoforms do not seem to be
3 maintained by higher reproductive fitness fo females that do not carry the thermostable allele
4 needed during the colonisation of new ‘hot’ chimneys. This persistence over a long period of
5 time may rather be explained by a thermodynamic trade-off between protein thermostability
6 and catalysis of these functional phenotypes.

7 8 Acknowledgments

9 This article has been written in memory of Dominique Le Guen, who greatly helped us in setting
10 up of the allozyme genotyping. We thank chief scientists and the ‘Nautile’ crews for their
11 technical support and efforts during the oceanographic expeditions Phare2002, Biospeedo2004
12 and Mescal2010. We are very grateful to the two anonymous referees who provided valuable
13 comments and editorial suggestions on the manuscript.

14 15 Literature Cited

- 16 [1] Holt, R. D., and M. S. Gaines, 1992 Analysis of adaptation in heterogeneous landscapes: implications
17 for the evolution of fundamental niches. *Evolut. Ecol.* 6: 433-447.
- 18 [2] Schmidt, P. S., E. A. Serrão, G. A. Pearson, C. Riginos, P. D. Rawson, *et al.*, 2008 Ecological
19 genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology* 89: 91-
20 107.
- 21 [3] Nevo, E., T. Shimony, and M. Libni, 1977 Thermal selection of allozyme polymorphisms in
22 barnacles. *Nature*, 267(5613): 699.
- 23 [4] Nevo, E., E. Lavie, and R. Ben-Shlomo, 1983 Selection of allelic isozyme polymorphisms in marine
24 organisms: pattern, theory, and application. *Isozymes*, 10: 69-92.
- 25 [5] Nevo, E., R. Noy, B. Lavie, A. Beiles, and S. Muchtar, 1986 Genetic diversity and resistance to
26 marine pollution. *Biol. J. Linn. Soc.* 29: 139-144.
- 27 [6] Pogson, G. H., 1989 Biochemical characterization of genotypes at the phosphoglucomutase-2 locus
28 in the Pacific oyster, *Crassostrea gigas*. *Biochem. Genetics*, 27: 571-589.
- 29 [7] Riddoch, B., 1993 The adaptive significance of electrophoretic mobility in phosphoglucose
30 isomerase (PGI). *Biol. J. Linn. Soc.* 50: 1-17.
- 31 [8] Schmidt, P. S., and D. M. Rand, 2001 Adaptive maintenance of genetic polymorphism in an intertidal
32 barnacle: habitat-and life-stage-specific survivorship of MPI genotypes. *Evolution* 55: 1336-1344.
- 33 [9] Boutet, I., A. Tanguy, D. Le Guen, P. Piccino, S. Hourdez, *et al.*, 2009 Global depression in gene
34 expression as a response to rapid thermal changes in vent mussels. *Proc. Roy. Soc. London B* 276: 3071-
35 3079.

- 1 [10] Bougerol, M., I. Boutet, D. Le Guen, D. Jollivet, and A. Tanguy, 2015 Transcriptomic response of
2 the hydrothermal mussel *Bathymodiolus azoricus* mediated by heavy metals is modulated by *Pgm*
3 genotypes and symbiont content. *Mar. Genomics*, <http://dx.doi.org/10.1016/j.margen.2014.11.010>.
- 4 [11] Eanes, W. F., 1999 Analysis of selection on enzyme polymorphisms. *Ann. Rev. Ecol. Syst.* 30:
5 301-326.
- 6 [12] Piccino, P., F. Viard, P.-M. Sarradin, N. Le Bris, D. Le Guen, *et al.*, 2004 Thermal selection of
7 PGM allozymes in newly founded populations of the thermotolerant vent polychaete *Alvinella*
8 *pompejana*. *Proc. Roy. Soc. London B* 271: 2351-2359.
- 9 [13] Pogson, G. H., 1991 Expression of overdominance for specific activity at the phosphoglucosmutase-
10 2 locus in the Pacific oyster, *Crassostrea gigas*. *Genetics*, 128: 133-141.
- 11 [14] Verrelli, B. C., and W. F. Eanes, 2001a Clinal variation for amino acid polymorphisms at the *Pgm*
12 locus in *Drosophila melanogaster*. *Genetics* 157: 1649-1663.
- 13 [15] Vrijenhoek, R. C., 1997 Gene flow and genetic diversity in naturally fragmented metapopulations
14 of deep-sea hydrothermal vent animals. *J. Hered.* 88: 285-293.
- 15 [16] Jollivet, D., P. Chevaldonné, B. Planque, 1999 Hydrothermal-vent alvinellid polychaete dispersal
16 in the Eastern Pacific. 2. A metapopulation model based on habitat shifts. *Evolution* 53: 1128-1142.
- 17 [17] Vrijenhoek, R. C., 2010 Genetic diversity and connectivity of deep-sea hydrothermal vent
18 metapopulations. *Mol. Ecol.* 19: 4391-4411.
- 19 [18] Watremez, P., and C. Kervevan, 1990 Origine des variations de l'activité hydrothermale: premiers
20 éléments de réponse d'un modèle numérique simple. *C. R. Acad. Sci Paris, Sér. 2*, 311: 153-158.
- 21 [19] Johnson, K. S., J. J. Childress, R. R. Hessler, C. M. Sakamoto-Arnold, and C. L. Beehler, 1988
22 Chemical and biological interactions in the Rose Garden hydrothermal vent field, Galapagos spreading
23 center. *Deep Sea Res. A* 35: 1723-1744.
- 24 [20] Le Bris, N., and F. Gaill, 2007 How does the annelid *Alvinella pompejana* deal with an extreme
25 hydrothermal environment? *Rev. Environ. Sci. Biotechnol.* 6: 197-221.
- 26 [21] Hourdez, S., and Lallier, F. H. 2006 Adaptations to hypoxia in hydrothermal-vent and cold-seep
27 invertebrates. In *Life in extreme environments*. Springer, Dordrecht, pp. 297-313.
- 28 [22] Von Damm, K. L., 1995 Controls on the chemistry and temporal variability of seafloor
29 hydrothermal fluids. pp. 222-247 in *Seafloor hydrothermal systems: Physical, chemical, biological, and*
30 *geological interactions*, 91.
- 31 [23] Jollivet, D., 1996 Specific and genetic diversity at deep-sea hydrothermal vents: an overview.
32 *Biodiv. Cons.* 5: 1619-1653.
- 33 [24] Matabos, M., N. Le Bris, S. Pendlebury, and E. Thiébaud, 2008 Role of physico-chemical
34 environment on gastropod assemblages at hydrothermal vents on the East Pacific Rise (13 N/EPR). *J.*
35 *Mar. Biol. Ass. UK* 88: 995-1008.

- 1 [25] Shank, T. M., D. J. Fornari, K. L. Von Damm, M. D. Lilley, R. M. Haymon, *et al.*, 1998 Temporal
2 and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9
3 50' N, East Pacific Rise). *Deep Sea Res. II: Topical Studies in Oceanography* 45: 465-515.
- 4 [26] Fontaine, F. J., M. Cannat, and J. Escartin, 2008 Hydrothermal circulation at slow-spreading mid-
5 ocean ridges: The role of along-axis variations in axial lithospheric thickness. *Geology* 36: 759-762.
- 6 [27] Marcus, J., V. Tunnicliffe, and D. A. Butterfield, 2009 Post-eruption succession of macrofaunal
7 communities at diffuse flow hydrothermal vents on Axial Volcano, Juan de Fuca Ridge, Northeast
8 Pacific. *Deep Sea Res. II: Topical Studies in Oceanography*, 56: 1586-1598.
- 9 [28] Chevaldonné, P., D. Desbruyères, and J. J. Childress, 1992 ... and some even hotter. *Nature*
10 359(6396): 593.
- 11 [29] Cary, S. C., T. Shank, and J. Stein, 1998 Worms bask in extreme temperatures. *Nature* 391(6667):
12 545.
- 13 [30] Ravaux, J., G. Hamel, M. Zbinden, A. A. Tasiemski, I. Boutet, *et al.*, 2013 Thermal limit for
14 metazoan life in question: *in vivo* heat tolerance of the Pompeii worm. *PLoS One*, 8(5): e64074.
- 15 [31] Jang, S. J., E. Park, W. K. Lee, S. B. Johnson, R. C. Vrijenhoek, *et al.*, 2016 Population subdivision
16 of hydrothermal vent polychaete *Alvinella pompejana* across equatorial and Easter Microplate
17 boundaries. *BMC Evolut. Biol.* 16: 235.
- 18 [32] Desbruyères, D., P. Chevaldonné, A.-M. Alayse, D. Jollivet, F. H. Lallier, *et al.*, 1998 Biology and
19 ecology of the “Pompeii worm”(*Alvinella pompejana* Desbruyères and Laubier), a normal dweller of an
20 extreme deep-sea environment: a synthesis of current knowledge and recent developments. *Deep Sea*
21 *Res. II: Topical Studies in Oceanography* 45: 383-422.
- 22 [33] Chevaldonné, P., C. R. Fisher, J. J. Childress, D. Desbruyères, D. Jollivet, *et al.*, 2000
23 Thermotolerance and the ‘Pompeii worms’. *Mar. Ecol. Prog. Ser.* 208: 293-295.
- 24 [34] Jollivet, D., D. Desbruyères, F. Bonhomme, and D. Moraga, 1995a Genetic differentiation of deep-
25 sea hydrothermal vent alvinellid populations (Annelida: Polychaeta) along the East Pacific Rise.
26 *Heredity*, 74: 376-391.
- 27 [35] Plouviez, S., D. Le Guen, O. Lecompte, F. H. Lallier, and D. Jollivet, 2010 Determining gene flow
28 and the influence of selection across the equatorial barrier of the East Pacific Rise in the tube-dwelling
29 polychaete *Alvinella pompejana*. *BMC Evol. Biol.*, 10: 220.
- 30 [36] Jollivet, D., D. Desbruyères, C. Ladrat, and L. Laubier, 1995b Evidence for differences in the
31 allozyme thermostability of deep-sea hydrothermal vent polychaetes (Alvinellidae): a possible selection
32 by habitat. *Mar. Ecol. Prog. Ser.* 123:125-136.
- 33 [37] Verrelli, B. C., and W. F. Eanes, 2001b The functional impact of *Pgm* amino acid polymorphism
34 on glycogen content in *Drosophila melanogaster*. *Genetics* 159: 201-210.
- 35 [38] Charlesworth, D., 2006 Balancing selection and its effects on sequences in nearby genome regions.
36 *PLoS Genetics* 2: 379-384.
- 37 [39] Hedrick, P. W., 2007 Balancing selection. *Current Biol.* 17: 230-231.

- 1 [40] Plouviez, S., T. M. Shank, B. Faure, C. Daguin-Thiébaud, F. Viard, *et al.*, 2009 Comparative
2 phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant
3 processes and population expansion in the South. *Mol. Ecol.* 18: 3903-3917.
- 4 [41] Pradillon, F., B. Shillito, C. M. Young, and F. Gaill, 2001 Deep-sea ecology: Developmental arrest
5 in vent worm embryos. *Nature*, 413(6857): 698.
- 6 [42] Faure, B., P. Chevaldonné, F. Pradillon, E. Thiébaud, E., and D. Jollivet, 2007 Spatial and temporal
7 dynamics of reproduction and settlement in the Pompeii worm *Alvinella pompejana* (Polychaeta:
8 Alvinellidae). *Mar. Ecol. Prog. Ser.* 348: 197-211.
- 9 [43] Mishra, R. N., S. L. Singla-Pareek, S. Nair, S. K. Sopory, and M. K. Reddy, 2002 Directional
10 genome walking using PCR. *Biotechniques*, 33: 830-842.
- 11 [44] Bierne, N., A. Tanguy, M. Faure, B. Faure, E. David, E., *et al.*, 2007 Mark-recapture cloning: a
12 straightforward and cost-effective cloning method for population genetics of single copy nuclear DNA
13 sequences in diploids. *Mol. Ecol. Notes* 7: 562-566.
- 14 [45] Black, W. C., and E. S. Krafur, 1985 A FORTRAN program for the calculation and analysis of
15 two-locus linkage disequilibrium coefficients. *Theor. Appl. Genet.* 70: 491-496.
- 16 [46] Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Catch, 2004 *GENETIX 4.0. 5.2, Software*
17 *under Windows™ for the genetics of the populations*. University of Montpellier, Montpellier, France.
- 18 [47] Bandelt, H. J., P. Forster, and A. Röhl, 1999 Median-joining networks for inferring intraspecific
19 phylogenies. *Mol. Biol. Evol.* 16: 37-48.
- 20 [48] Librado, P., and J. Rozas, 2009 DnaSP v5: a software for comprehensive analysis of DNA
21 polymorphism data. *Bioinformatics*, 25: 1451-1452.
- 22 [49] Kelly, J. K., 1997 A test of neutrality based on interlocus associations. *Genetics*, 146: 1197-1206.
- 23 [50] Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events
24 in the history of a sample of DNA sequences. *Genetics* 111: 147-164.
- 25 [51] Martin, D. P., P. Lemey, M. Lott, V. Moulton, D. Posada, *et al.*, 2010 RDP3: a flexible and fast
26 computer program for analyzing recombination. *Bioinformatics*, 26: 2462-2463.
- 27 [52] Li, N. and M. Stephens, 2003 Modelling linkage disequilibrium and identifying recombination
28 hotspots using snp data. *Genetics* 165, 2213-2233.
- 29 [53] Hudson, R. R., M. Slatkin, and W. P. Maddison, 1992 Estimation of levels of gene flow from DNA
30 sequence data. *Genetics* 132: 583-589.
- 31 [54] Kumar, S., G. Stecher, and K. Tamura, 2016 MEGA7: molecular evolutionary genetics analysis
32 version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33: 1870-1874.
- 33 [55] Ewing, G., and J. Hermisson, 2010 MSMS: a coalescent simulation program including
34 recombination, demographic structure and selection at a single locus. *Bioinformatics*, 26(16): 2064-
35 2065.
- 36 [56] Thornton K. 2003 Libsequence: a c++ class library for evolutionary genetic
37 analysis. *Bioinformatics*, 19(17) : 2325-2327.

- 1 [57] Reikofski, J., and B. Y. Tao, 1992 Polymerase chain reaction (PCR) techniques for site-directed
2 mutagenesis. *Biotechnol. Adv.* 10: 535-547.
- 3 [58] Pace, C. N., and J. M. Scholtz, 1997 Measuring the conformational stability of a protein. *Protein*
4 *structure: A practical approach*, 2: 299-321.
- 5 [59] Pace, C. N., 1986 Determination and analysis of urea and guanidine hydrochloride denaturation
6 curves. pp. 266-280 in *Methods in enzymology* Vol. 131, Academic Press.
- 7 [60] Walter, P., and D. Ron, 2011 The unfolded protein response: from stress pathway to homeostatic
8 regulation. *Science* 334: 1081-1086.
- 9 [61] Sali, A., and T. L. Blundell, 1993 Comparative protein modelling by satisfaction of spatial
10 restraints. *J. Mol. Biol.* 234: 779-815.
- 11 [62] Guerois, R., J. E. Nielsen, and L. Serrano, 2002 Predicting changes in the stability of
12 proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* 320: 369-387.
- 13 [63] Whitehouse, D. B., W. Putt, J. U. Lovegrove, K. Morrison, M. Hollyoake, *et al.*, 1992
14 Phosphoglucomutase 1: complete human and rabbit mRNA sequences and direct mapping of this highly
15 polymorphic marker on human chromosome 1. *Proc. Nat. Acad. Sci. USA* 89: 411-415.
- 16 [64] Tanguy, A., I. Boutet, P. Boudry, L. Degremont, J. Laroche, *et al.*, 2006 Molecular identification
17 and expression of the phosphoglucomutase (PGM) gene from the Pacific oyster *Crassostrea gigas*.
18 *Gene*, 382: 20-27.
- 19 [65] Matabos M., S. Plouviez, S. Hourdez, D. Desbruyères, P. Legendre, *et al.*, 2011 Faunal changes
20 and geographic crypticism indicate the occurrence of a biogeographic transition zone along the Southern
21 East-Pacific Rise. *J. Biogeogr.* 38: 575-594.
- 22 [66] Matabos, M., and D. Jollivet, 2019 Revisiting the species' complex of *Lepetodrilus elevatus*
23 (Vetigastropod, Lepetodrilidae) using gastropod samples from the Galápagos and Guaymas
24 hydrothermal vent systems. *J. Moll. Stud.* 85: 154-165.
- 25 [67] Eanes, W. F., 2011 Molecular population genetics and selection in the glycolytic pathway. *J. Exp.*
26 *Biol.* 214: 165-171.
- 27 [68] Verrelli, B. C., and W. F. Eanes, 2000 Extensive amino acid polymorphism at the *Pgm* locus is
28 consistent with adaptive protein evolution in *Drosophila melanogaster*. *Genetics*, 156: 1737-1752.
- 29 [69] Wheat, C. W., W. B. Watt, D. D. Pollock, and P. M. Schulte, 2005 From DNA to fitness differences:
30 sequences and structures of adaptive variants of *Colias* phosphoglucose isomerase (PGI). *Mol. Biol.*
31 *Evol.* 23: 499-512.
- 32 [70] McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus in
33 *Drosophila*. *Nature*, 351(6328), 652-654.
- 34 [71] Begun, D. J., Betancourt, A. J., Langley, C. H., & Stephan, W. (1999). Is the fast/slow allozyme
35 variation at the *Adh* locus of *Drosophila melanogaster* an ancient balanced polymorphism?. *Molecular*
36 *Biology and Evolution*, 16(12), 1816-1819.

- 1 [72] Fontanillas, E., O.V. Galzitskaya, O. Lecompte, M. Y. Lobanov, A. Tanguy, *et al.*, 2017 Proteome
2 evolution of deep-sea hydrothermal vent alvinellid polychaetes supports the ancestry of thermophily
3 and subsequent adaptation to the cold for some lineages. *Genome Biol. Evol.* 9: 279-296.
- 4 [73] Bioy, A., 2018 *Histoire évolutive et influence de la sélection sur la diversité génétique des annélides*
5 *polychètes d'environnements extrêmes*. Thèse de Doctorat, Sorbonne Université, 248 pp.
- 6 [74] Hoekstra, R. F., R. Bijlsma, and A. J. Dolman, 1985 Polymorphism from environmental
7 heterogeneity: models are only robust if the heterozygote is close in fitness to the favoured homozygote
8 in each environment. *Genet. Res.* 45: 299-314.
- 9 [75] Gardner, J. P. A., and I. Lobkovy, 2005 A test for overdominance at the phosphoglucosylase-2 locus
10 in Pacific oysters (*Crassostrea gigas*) from B-New Zealand. *Aquaculture* 244: 29-39.
- 11 [76] Levene, H., 1953 Genetic equilibrium when more than one ecological niche is available. *Am. Nat.*
12 87: 331-333.
- 13 [77] Gillespie, J. H., 1985 The interaction of genetic drift and mutation with selection in a fluctuating
14 environment. *Theor. Pop. Biol.* 27: 222-237.
- 15 [78] Hedrick, P. W., M. E. Ginevan, and E. P. Ewing, 1976 Genetic polymorphism in heterogeneous
16 environments. *Ann Rev Ecol. Syst.* 7: 1-32.
- 17 [79] Hedrick, P. W., 1986 Genetic polymorphism in heterogeneous environments: a decade later. *Ann.*
18 *Rev. Ecol. Syst.* 17: 535-566.
- 19 [80] Watt, W. B., 1977 Adaptation at specific loci. I. Natural selection on phosphoglucose isomerase of
20 *Colias* butterflies: biochemical and population aspects. *Genetics* 87: 177-194.
- 21 [81] Watt, W. B., R. C Cassin, and M. S. Swan, 1983 Adaptation at specific loci. III. Field behavior and
22 survivorship differences among *Colias* PGI genotypes are predictable from *in vitro* biochemistry.
23 *Genetics* 103: 725-739.
- 24 [82] Carter, P. A., and W. B. Watt, 1988 Adaptation at specific loci. V. Metabolically adjacent enzyme
25 loci may have very distinct experiences of selective pressures. *Genetics* 119: 913-924.
- 26 [83] Hartl, D. L., D. E. Dykhuizen, and A. M. Dean, 1985 Limits of adaptation: the evolution of selective
27 neutrality. *Genetics*, 111: 655-674.
- 28 [84] Somero, G. N., 1978 Temperature adaptation of enzymes: biological optimization through
29 structure-function compromises. *Ann. Rev. Ecol. Syst.* 9: 1-29.
- 30 [85] Somero, G. N., 1995 Proteins and temperature. *Ann. Rev. Physiol.* 57: 43-68.
- 31 [86] Vogt, G., S. Woell, P. Argos, 1997 Protein thermal stability, hydrogen bonds, and ion pairs. *J. Mol.*
32 *Biol.* 269: 631-643.
- 33 [87] Shin, D. S., M. DiDonato, D. P. Barondeau, G. L. Hura, C. Hitomi, *et al.*, 2009 Superoxide
34 dismutase from the eukaryotic thermophile *Alvinella pompejana*: structures, stability, mechanism, and
35 insights into amyotrophic lateral sclerosis. *J. Mol. Biol.* 385: 1534-1555.
- 36 [88] Bruneaux, M., J. Mary, M. Verheye, O. Lecompte, O. Poch, *et al.*, 2013 Detection and
37 characterisation of mutations responsible for allele-specific protein thermostabilities at the Mn-

- 1 superoxide dismutase gene in the deep-sea hydrothermal vent polychaete *Alvinella pompejana*. J. Mol.
- 2 Evol. 76: 295-310.

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

3

Mutation	D_{ij}	R_{ij}	K_{hi}^2	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*

4

1 Table 2. Gene diversities, population parameters and neutrality tests along the *Pgm-1* gene for *A. pompejana* populations of the South and North EPR. N
2 and S represent the number of sequences and the number of segregating sites used, respectively. Linkage disequilibria between sites were only estimated
3 between informative sites only: numbers in brackets correspond to the number of significant exact Fisher tests, total number of comparisons and numbers
4 of tests still significant after the Bonferonni correction, respectively. (RDP n.d.): recombinant not detected using automated RDP and bootscan packages
5 of RDP v.3.44. Values in brackets below π_S and π_N (Jukes & Cantor estimates) are the numbers of synonymous and non-synonymous sites in coding
6 regions, respectively. All genetic datasets obtained using the MCR method were corrected for artifactual/somatic singletons.

7

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 _d	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 π	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
π_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)
π_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 _{nS}	0.054 (1/15/1 ^B)	0.008 (0/1/0 ^B)	0.110 (43/325/19 ^B)	0.466 (46/153/0 ^B)	0.0093 (2/36/1 ^B)	0.0540 (11/66/7 ^B)	0.0226 (0/3/0 ^B)	0.0324 (3/21/0 ^B)	0.0261 (4/120/1 ^B)	0.1641 (5/36/0 ^B)
R _m	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 _{st}	0.256***		0.262***		0.510***		0.291**		0.015*	
D _{xy}	0.0051		0.0098		0.0117		0.0059		0.0059	
Tajima's D	-0.46 ^{NS}	-1.52 ^{NS}	-0.07 ^{NS}	+0.46 ^{NS}	-1.57 ^{NS}	-1.12 ^{NS}	-1.07 ^{NS}	-0.66 ^{NS}	-1.73 ^{NS}	-0.41 ^{NS}
Fu & Li's F	-0.19 ^{NS}	-1.89 ^{NS}	+0.07 ^{NS}	+0.34 ^{NS}	-2.31*	-1.15 ^{NS}	-0.69 ^{NS}	-0.57 ^{NS}	-1.40 ^{NS}	+0.07 ^{NS}

1 ^B: still significant after a Bonferonni test. Level of significance following permutation tests (1000 re-samplings): * <0.05, **<0.01, ***<0.001, ^{NS}: not
2 significant.

3
4
5

1 Table 3. Conformational and temperature stability of the three overexpressed variants (PGM78, PGM90, and PGM100). C_m et m values estimated
 2 from the variation of protein fluorescence in presence of an increasing concentration of GdmHCl (values for each of the two transitions). Estimation
 3 of the free enthalpy of the unfolding reaction in absence of chaotropic agent for each of the two transition states. T_m : values of the temperature at
 4 which we reach 50% of non-reversible inactivation after a 30-minute exposure. K_m^{app} and K_{cat} are kinetic parameters corresponding to the apparent
 5 Michaelis-Menten constant for glucose-1-phosphate, and the catalytic constant, respectively. The ratio of these two values corresponds to the
 6 specific activity.

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

9

Figure captions

Fig. 1. Species range of the Pompeii worm *Alvinella pompejana* along the East Pacific Rise. Dashed line indicates the presence of the Equatorial barrier to gene flow depicted by Plouviez et al. (2009, 2010[35,44]). Blue and Red boxes correspond to the northern and southern metapopulations of the worm.

Fig. 2. Map of the *A. pompejana Pgm-1* gene with the human (*Homo sapiens*) and the oyster (*Crassostrea gigas*) PGM as comparison. Identification of the distinct loci sequenced with the method used (Mark, Cloning, Recapture (MCR) or direct sequencing) and population origin.

Fig. 3. Evolution of gene diversity (π) and the statistic Tajima's D along the *AP-Pgm-1* gene using a sliding window of 100 bp size and a step of 25 bp. The analysis includes exonic and intronic fragments for which the sequence polymorphism has been documented. Arrows indicate the portions of the gene for which there are no genetic dataset.

Fig. 4. Minimum evolution tree obtained from evolutionary distances computed with the Maximum Composite Likelihood method with MEGA7 on 48 sequences from individuals of the North and South EPR locations using the Mark-Cloning-Recapture (MCR) of the *Pgm-1* introns 2 and exon 3 (1110 bp). The sequences corresponding to the PGM 78, 90 and 100 are respectively identified by the letters QE, EQ and EE traducing the polymorphism at positions 155 and 190, with the colours blue and red corresponding respectively to the individuals from the north and the south.

Fig. 5. Residual enzyme activities after 30 min of incubation at different temperatures for the overexpressed isoforms of PGM 78 (QE), 90 (EQ) and 100 (EE). T_m values are shown in Table 4.

Fig. 6. 3D structural model of *A. pompejana* PGM 78 fitted on the PGM-1 rabbit template (1C47, 2.70Å) using Modeller 9v13. The protein is structured in 4 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 to catalytic site of the enzyme, which binds the reaction catalyser, alpha-D-glucose-1,6-diphosphate, and the ion Mg^{2+} .

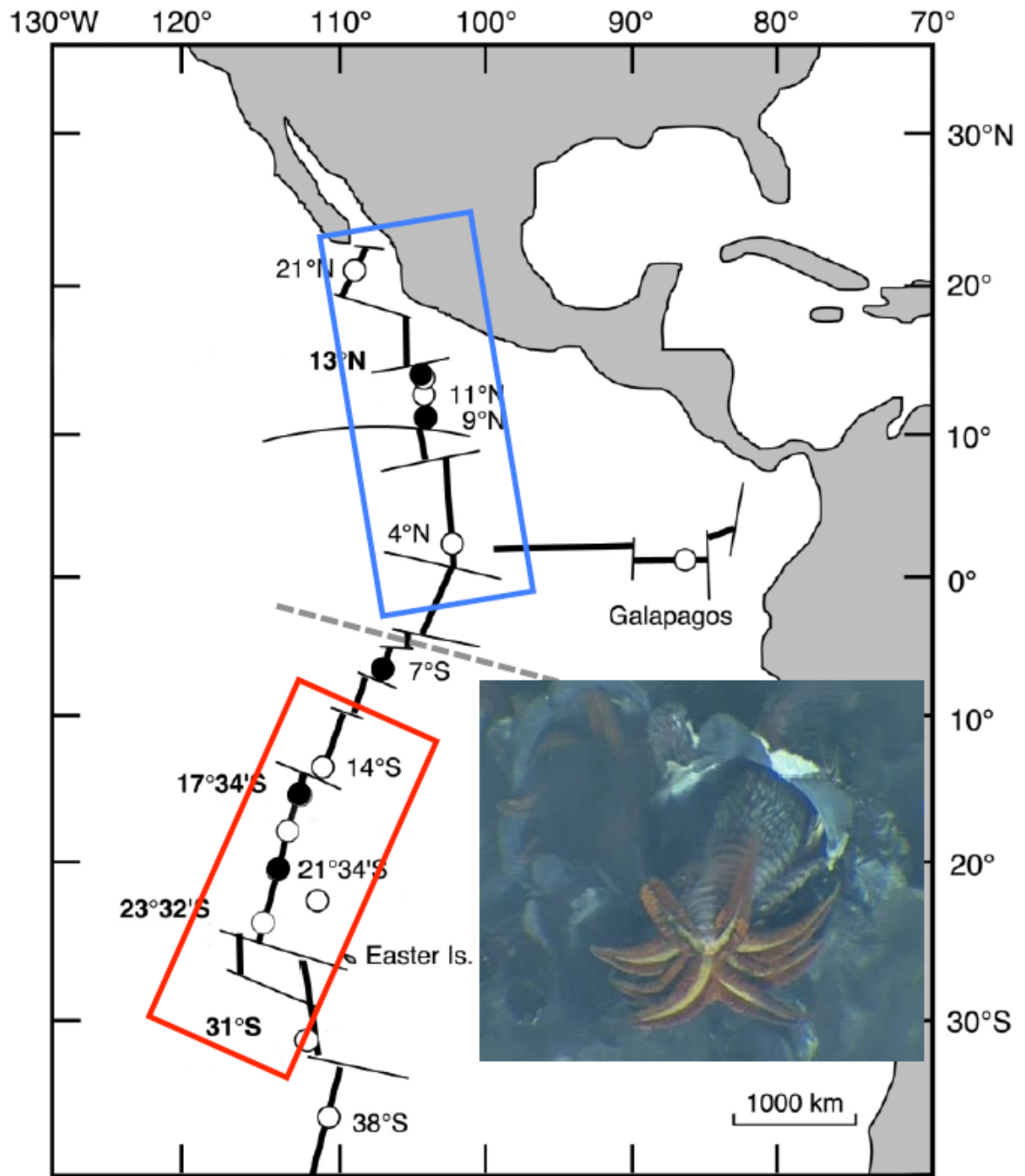


Fig. 1

Fig. 2

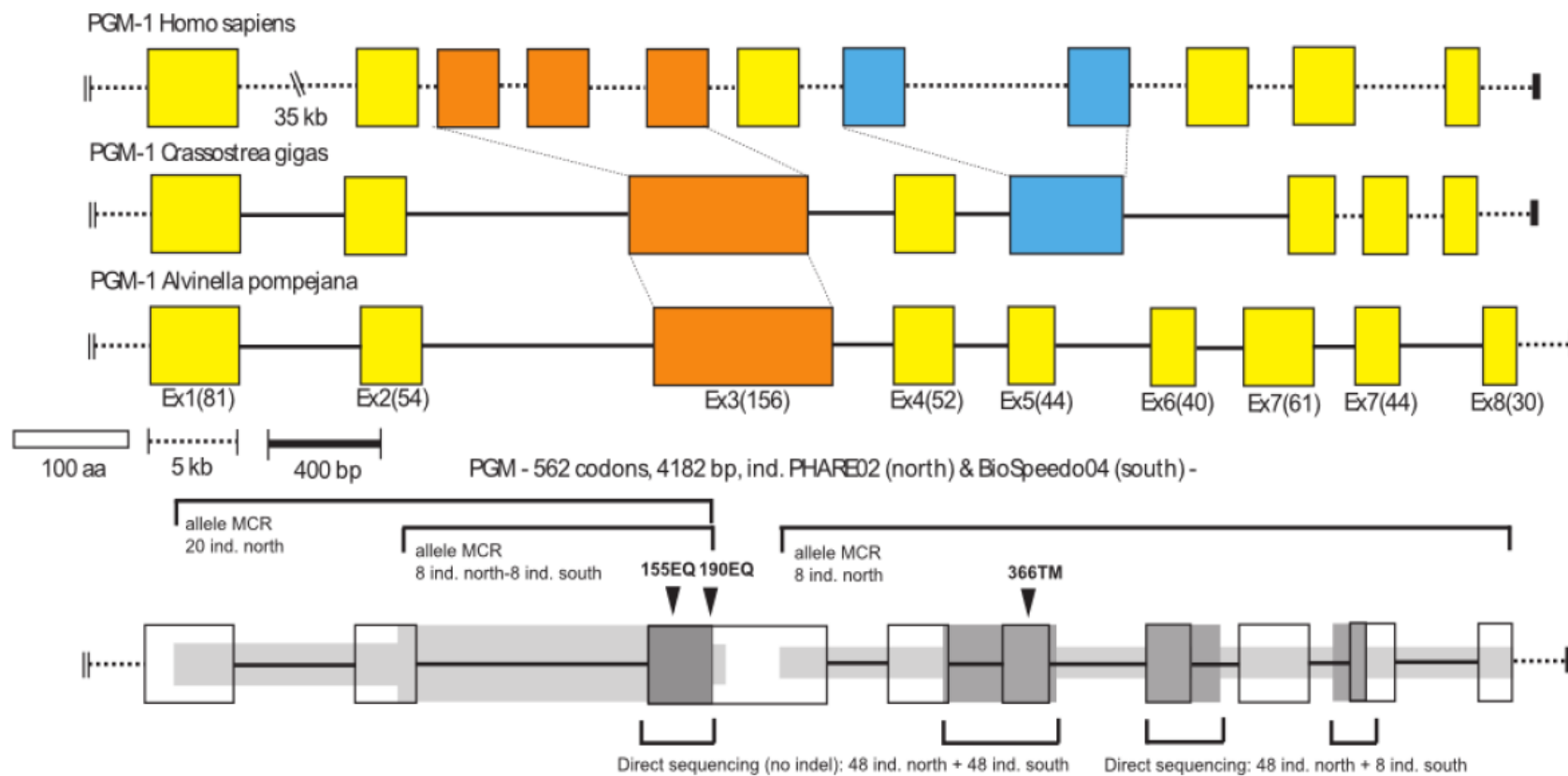


Fig. 3

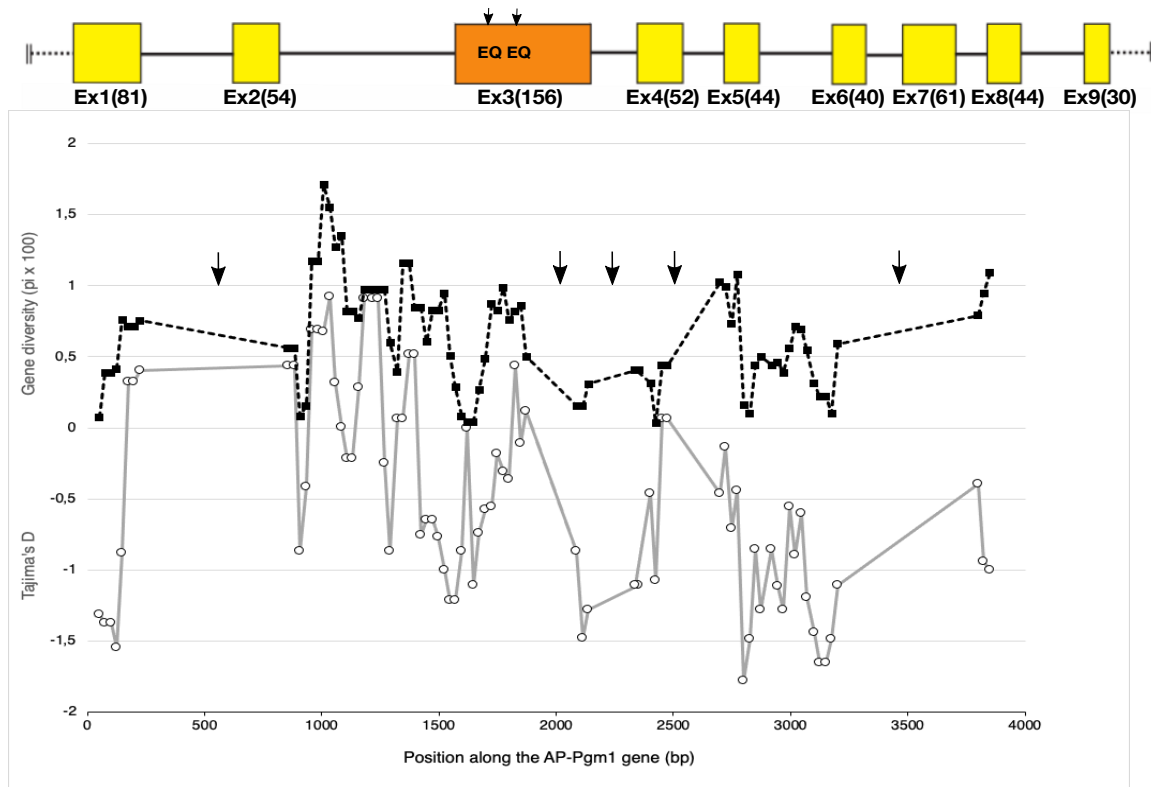


Fig. 5

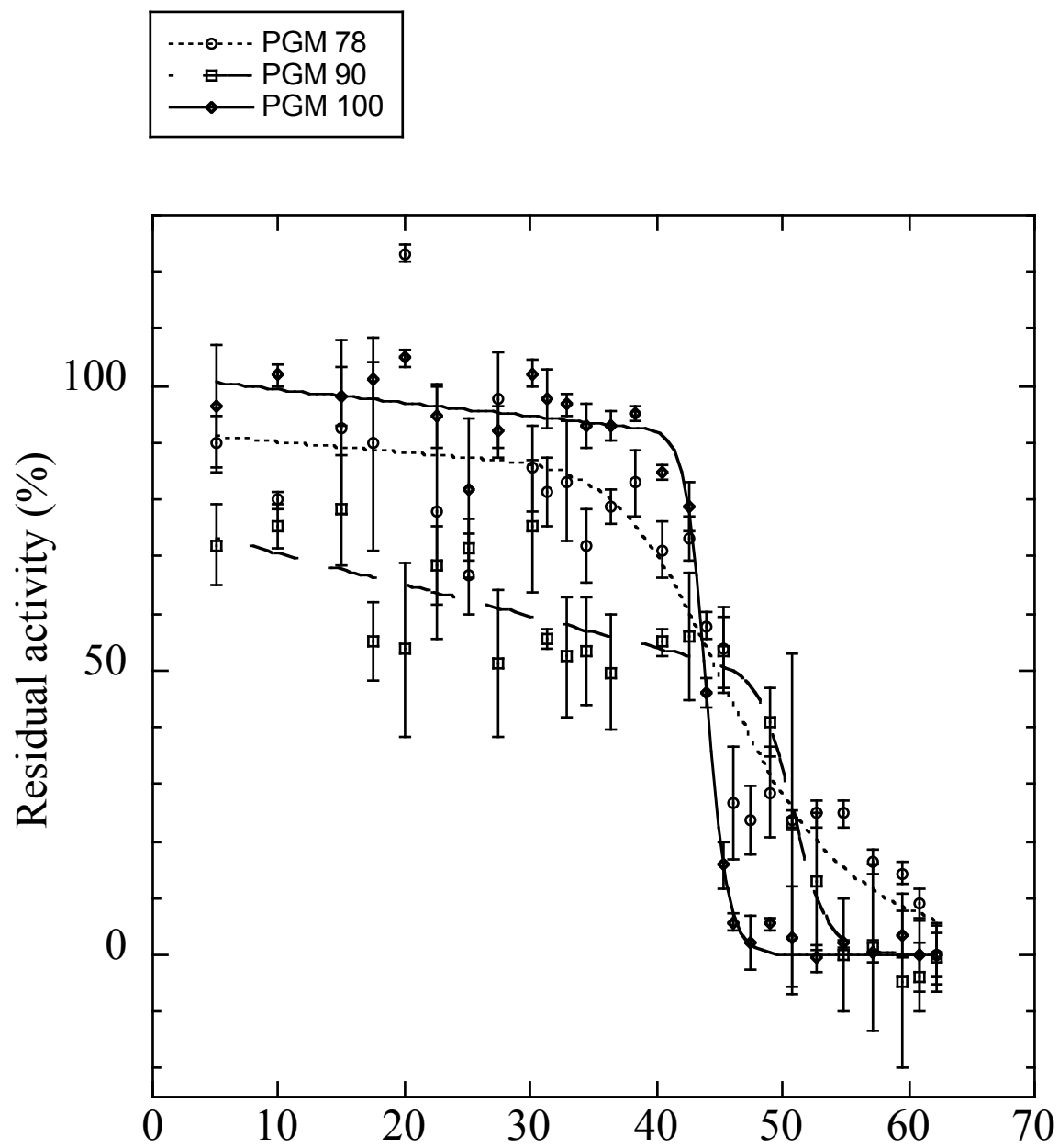


Fig. 6

