

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.

13 Email: 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
19 Banyuls, 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 and, display specific adaptations to
3 withstand high temperature and hypoxia associated with this highly variable habitat. Previous
4 studies revealed the existence of a balanced polymorphism on the enzyme
5 phosphoglucomutase associated with thermal variations where allozymes 90 and 100
6 exhibited different optimal activities and thermostabilities. The exploration of the mutational
7 landscape of the phosphoglucomutase1 revealed the maintenance of four highly divergent
8 allelic lineages encoding the three most frequent electromorphs over the worm's geographic
9 range. This polymorphism is only governed by two linked amino-acid replacements located in
10 exon 3 (E155Q and E190Q). A two-niches model of selection with 'cold' and 'hot' conditions
11 represents the most likely way for the long-term persistence of these isoforms. Using directed
12 mutagenesis, overexpression of the three recombinant variants allowed us to test the additive
13 effect of these two mutations on the biochemical properties of this enzyme. Results are
14 coherent with those previously obtained from native proteins and reveal a thermodynamic
15 trade-off between the protein thermostability and catalysis, which is likely to have maintained
16 these functional phenotypes prior to the geographic separation of populations across the
17 Equator, about 1.2 Mya.

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 phenotype
5 of a population evolves under heterogeneous and/or changing conditions^[1]. As a consequence,
6 many studies have investigated the maintenance of enzyme polymorphism 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
9 mostly associated with isomerase and mutase functions, such as GPI, MPI or PGM, have been
10 shown to display isoforms that may be the subject of natural selection, leading to habitat-
11 driven differentiation in populations according to temperature, wave action or metallic
12 pollution^[2,3,4,5,6,7,8,9,10]. According to Eanes^[11], such branch-point enzymes, which are
13 positioned at the crossroad of metabolic pathways, are likely to be the target of natural
14 selection as they can orientate pathway flux according to their protein variation. Amongst
15 them, alleles encoding the enzyme phosphoglucomutase have been widely studied with the
16 aim of testing the hypothesis of differential and/or balancing selection by looking at
17 allele^[3,4,12], and heterozygote^[13] frequencies in populations but also to assess either the fitness
18 of individuals carrying alleles suspected to be advantageous along latitudinal clines^[14] or the
19 kinetic properties of the enzyme isoforms themselves^[6,12].

20 Because of their tremendous thermal variability due to the chaotic mixing of the cold
21 sea water and hot fluids, hydrothermal vents represent an ideal model for testing the effect of
22 frequent -and unpredictable- spatial and temporal changes of habitats on 'adaptive' enzyme
23 polymorphisms. First, both the fragmentation and instability of the vent discharge likely
24 promote highly dynamic meta-populations with recurrent local extinctions and associated
25 bottlenecks^[15,16,17]. Long-term oscillations of the heat convection beneath the ridge lead to the

1 displacement of the hydrothermal activity along the rift, generating the emergence of new
2 vent sites more or less closely to older ones that became extinct^[16,18]. Second, variations of
3 temperature, sulphide and oxygen concentrations over short periods of time (often ranging
4 from minutes to hours^[19,20] are likely to affect the respiratory, nutritional and reproductive
5 physiologies of animals living there^[21]. Such a constant variability of the vent conditions at
6 the individual scale represents a selective constraint that should ever promotes the emergence
7 of highly plastic allelic isoforms or the maintenance of highly specialized alleles by favouring
8 the heterozygous state. Finally, the hydrothermal environment is highly fragmented and
9 heterogeneous according to the mineral composition of the oceanic crust through which the
10 super-heated fluid is moving prior to be expelled above the seafloor^[22]. As a consequence,
11 vent fields often correspond to a mosaic of edifices of different ages^[23,24], whose mean age
12 and size distribution is dictated by the frequency of tectonic and volcanic events and the
13 dynamics of the heat convection beneath oceanic ridges^[25,26,27]. This provides distinct
14 ecological niches for a variety of vent species able to colonise chimneys over time and thus an
15 ecological basis for diversifying selection.

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 hereafter) to 38°S on the Pacific Antarctic Ridge (PAR)^[31] and has developed
20 peculiar physiological adaptations to colonize this hostile habitat^[32,33]. Earlier genetic studies
21 showed that *A. pompejana* exhibits quite an unusual high level of genetic diversity^[31,34,35]
22 with a non-negligible number of bi-allelic enzyme loci with equiprequent alleles, some of
23 which displaying different thermal stabilities^[36]. Amongst them, the enzyme
24 phosphoglucomutase (PGM-1) possesses four distinct isoforms. Allozymes 90 and 100 have
25 frequencies of approximately 35% and 60%, respectively in populations of the Northern EPR,

1 the two other isoforms (112 and 78) rather rare, account for the remaining 5%. Although the
2 frequency of allozyme 90 remained constant over the species range, Plouviez *et al.*^[35]
3 however showed that allozymes 78 and 100 displayed an abrupt clinal distribution across the
4 Equator with allozyme 78 becoming the most frequent allele in the Southern EPR. Bi-allelism
5 was thus preserved all along the EPR despite population isolation, recurrent
6 extinction/recolonizations and a long history of divergence across the Equatorial barrier.

7 In addition, significant genetic differentiation has been observed between the worm
8 populations living in contrasted microhabitats, and especially when comparing newly formed
9 ‘still hot’ chimneys (‘hot’ niche) to older and colder edifices (‘cold’ niche). The frequency of
10 allele 90 was indeed positively correlated with mean temperature at the opening of *Alvinella*
11 tubes and increased in the former habitat suggesting that this locus was under diversifying
12 selection^[12]. *In vitro* experiments on the enzyme stability and optimum strengthened this view
13 and showed that allele 90 was more thermostable and more active at higher temperatures than
14 allele 100, and thus favoured in the ‘hot’ habitat but not in the ‘cold’ one.

15

16 Although whole-length PGM sequences have now been obtained for a large panel of
17 metazoan species, very few studies have been conducted at the population level, and most of
18 them involved bacterial strains. While this enzyme has been extensively studied in 1970-
19 1990’s for adaptive purposes, only few studies examined the relationship between non-
20 synonymous changes at the gene level and the subsequent enzyme performance of the
21 isoforms (but see^[14,37]) for the correspondence between allelism, enzyme thermal resistance
22 and glycogen storage in *Drosophila*). Here, we report a possible case of long-term balancing
23 selection at an enzyme locus where alleles can be maintained by a two-niche model of
24 selection where the proportions of the two niches greatly vary over space and time. Most of
25 the documented cases for the long-term persistence of alleles by balancing selection, and

1 trans-species polymorphism come from studies dealing with negative frequency-dependent
2 selection at immune and sex-determination genes^[38,39]. This raises questions about how a
3 chaotic and highly fluctuating two-niche system can promote balancing selection at key
4 branch-point enzymes.

5

6 MATERIAL and METHODS

7

8 Animal sampling

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

15 In order to test an earlier hypothesis^[12] postulating that individuals carrying genotypes
16 favored during the colonization of newly-formed still 'hot' chimneys may be counter-selected
17 by a lower reproductive fitness under cooler conditions (i.e. trade-off between settlement
18 ability and reproduction), we examined the relationship between PGM-1 genotypes and
19 female's fecundity. To this extent, the size of animals was estimated from the width at the S4
20 setigerous segment and sexes were determined based on the presence of either a genital pore
21 in females or a pair of sexual tentacles in males^[40]. Mature females collected from both sides
22 of the Equator were dissected to estimate their fecundity per size unit and genotyped at the
23 PGM-1 locus. For each female, the coelomic fluid containing oocytes was carefully removed
24 and resuspended in 50 ml of a solution of borate-buffered 3% formalin in seawater. Oocytes
25 were counted following the method previously described by Faure *et al.*^[41] and a one-way

1 ANOVA was performed on size-corrected female fecundities according to the genotype using
2 the software Jamovi (<https://www.jamovi.org>).

3

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

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

6 Based on allozyme genotypes, 8 homozygous individuals carrying alleles 78, 90 and
7 100 were selected for RNA extractions. Total RNAs were extracted with Tri-Reagent (Sigma)
8 following the manufacturer's instructions and a classical chloroform extraction protocol. Both
9 the quantity and quality of RNAs were assessed with a Nanodrop ND-1000
10 spectrophotometer (Nanodrop Technologies, Delaware, USA). Five μ g of total RNAs were
11 reverse transcribed with a M-MLV reverse transcriptase (Promega), an anchor-oligo(dT)
12 primer (Table S1) and random hexamers (Promega). The reverse anchor and forward nested
13 degenerated PGM primers derived from the oyster *Crassostrea gigas* and human *Pgm-1*
14 sequences were then used to perform the upstream amplification of cDNA fragments (see
15 Table S1). PCR-products containing *Pgm-1* cDNA candidates were then cloned with the
16 TOPO TA Cloning kit (Invitrogen) and, sequenced on an ABI 3130 sequencer using the
17 BigDye v.3.1 (PerkinElmer) terminator chemistry following the manufacturer's protocol.
18 Sequences of clones containing the appropriate *Pgm-1* cDNA fragments were then aligned to
19 reconstruct a series of nearly complete *AP-Pgm-1* cDNA (i.e. only lacking a small part of the
20 5'end of the coding sequence).

21

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

23 Using gDNA, specific reverse primers (Table S1) were also used to amplify the 5'
24 portion of the gene by directional genome walking using PCR^[42]. A series of specific primers
25 were designed based on our cDNA sequences (see Table S1) to amplify both exon and intron-

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

16

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

18 To examine the correspondence between the only two diagnostic polymorphic non-
19 synonymous EQ mutations found at exon 3 and allozymes 78, 90 and 100, a total of 220
20 individuals were genotyped on the 350bp fragment of the *Pgm-1* exon 3 containing these
21 sites. PGM-1 allozymes were first screened for each individual by electrophoresis on 12%
22 starch-gel at 4°C (100 V, 80 mA, 4 h) with the Tris-citrate pH 8.0 buffer system following the
23 procedure described by Piccino^[12]. The 350 bp exon3 fragment was then amplified by PCR
24 on the same individuals following a gDNA extraction using a CTAB/PVP procedure
25 described by Plouviez^[44]. PCR amplifications were conducted using a specific primer pair

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

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

17

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

19 Nucleotidic diversities were punctually assessed along the gene by combining direct
20 sequencing and the MCR method on individuals from each side of the EPR (see Fig. 2). These
21 regions included exon 1, exons 4 to 5, end of exon 7 and the beginning of exon 9 (Accession
22 N° MN218840-MN218917 for exon1, Accession N° MN219292-MN219356 for exons 4 and
23 5). In addition, a fragment containing the whole intron 2 and the beginning of exon 3 where
24 the two diagnostic EQ mutations are located (1110 bp) was also sequenced using the MCR
25 method^[43] in order to test whether ‘hot spots’ of mutations occur around these two EQ sites

1 but also to estimate allele divergences (Accession N° MN219357 - MN219404). In the MCR
2 sequence sets, the number of retrieved alleles greatly varied between the different parts of the
3 gene according to the sequencing efficiency and/or cloning success. Artifactual singletons due
4 to the MCR method were removed by comparing the singleton rates between the MCR and
5 direct sequencing datasets on the same fragments.

6 Allele diversity (H_d), nucleotide diversity (π) and its synonymous and non-
7 synonymous components (π_S and π_N), and Watterson's theta (θ_w) were thus examined
8 together with deviations to neutral evolution (Tajima's D and Fu & Li's F statistics) for both
9 the northern and southern EPR individuals along the gene with the software DNAsp 4.10.3^[48]
10 using a sliding window (size=100 and step=50). These basic genetic parameters were then
11 compared with the critical values associated with sample size and neutral coalescent
12 simulations implemented in the same software. Linkage disequilibrium between segregating
13 sites and recombination among alleles were estimated by calculating the ZnS statistics^[49]
14 together with the minimum number of recombination events (Rm ^[50]). The number of
15 significant associations between linked sites was evaluated following a Fisher's exact test and
16 a Bonferroni correction implemented in DNAsp 4.10.3. The occurrence of recombinants was
17 also checked using automated RDP and bootscan packages of RDP v.3.44^[51] and the search
18 for hotspots in the recombination rate ($4N.r$) was examined along the gene (-recomb and -
19 hotspot outputs) for both the northern and southern populations using the software Phase
20 2.1.1^[52]. Genetic differentiation and allele divergence between the southern and northern parts
21 of EPR were estimated by calculating F_{st} and D_{xy} with DNAsp 4.10.3. Genetic
22 differentiation was tested using 1000 permutations of the sequence datasets using the
23 randomization test developed by Hudson^[53]. Finally, the intron 2-exon 3 alignment (1110 bp)
24 was used to reconstruct a coalescent tree of *AP-Pgm-1* alleles in order to evaluate more
25 specifically both the intra-locus recombination and allele divergence near the two EQ non-

1 synonymous polymorphic sites. The evolutionary history of alleles was inferred using the
2 Minimum Evolution method implemented in MEGA7^[54] using the NJ algorithm for the initial
3 tree, pairwise deletion of ambiguous sites, and the close-neighbor-interchange (CNI)
4 algorithm. Evolutionary distances were computed using the Maximum Composite Likelihood
5 method.

6

7 Coalescent simulations using models of selection

8 Additional simulations of a structured coalescent (N=1000 simulations with Ne=50
9 000) were also performed using the software msms v3.2^[55] with an asymmetrical migration
10 rate between two populations (pop_{1->2}: 2N.m=1 and pop_{2->1}: 2N.m=0.1) in a neutral way and
11 two different hypotheses of balancing selection: (1) overdominance with SaA=500 and
12 SAA=1, and (2) a 2-niches model of selection (4 populations and 2 habitats with SAA=1000,
13 SaA=500 and Saa=0 in the first habitat and SAA=0, SaA=500 and Saa=1000 in the second
14 habitat). Each set of simulations including the null hypothesis of asymmetrical migration
15 without selection were run with two recombination rates (R=1 or 100). Population parameters
16 including gene diversities (θ_w , π), and Tajima'D within each deme and Fst between demes
17 were estimated with the pylibseq 0.2.3 libraries^[56] using a home-made python script done by
18 this latter author.

19

20 Functional and structural analysis of PGM-1 recombinant isoforms

21 *Plasmid construction for enzyme overexpression*

22 Full-length AP-Pgm cDNA were obtained from, two homozygous individuals 100/100
23 (EE) and 90/90 (EQ). RT-PCR was conducted with the ClonTech SMARTer Race cDNA
24 amplification kit following the manufacturer instructions and AP-PGMex11 reverse primer
25 (see Table S1). These cDNAs were then used as a target to specifically amplify the complete

1 coding sequence with primers containing cutting sites to be inserted in either Pet20b or
2 PetDuet expression vectors (Table S1). Amplified coding sequences were double-digested
3 with either enzymes *BamHI/NotI* or *AseI/XhoI* in a 25 μ l volume containing the restriction
4 buffer, the enzymes, and 1% BSA. The restriction products were then ligated in the
5 appropriate expression vector after purification with a Nucleospin Gel extraction Clean up
6 column (Macherey Nagel) and cloned into BL21DE3 *E. coli* cells amenable for IPTG
7 induction and overexpression.

8

9 *Directed mutagenesis*

10 Using the full-length cDNA with the double mutation EE as a template, mutants ₁₅₅EQ
11 and ₁₉₀EQ were produced by directed mutagenesis following the PCR protocol of Reikofski
12 and Tao^[57]. First amplifications were conducted in 50 μ l reaction volume containing: 1X Pfu
13 buffer containing MgCl₂, 0.25 mM of each dNTP, 0.5 μ M of each primer (petDuet and
14 mutated primer), 0.5 U of the proof-reading *Pfu* polymerase (Promega) with 30 cycles of
15 94°C for 30s, 60°C for 30s and 72°C for 3 min. Secondly, the two regions of the mutated
16 cDNA were joined following a PCR amplification without primers mixing the two previous
17 PCR-products (1:1) under the same conditions and a final elongation step of 10 min. cDNAs
18 containing the mutated sites ₁₅₅E->Q and ₁₉₀E->Q and the native ₁₅₅E₁₉₀E cDNA were then
19 sequenced on an ABI 3130 sequencer with the BigDye v.3.1 (Perkin Elmer) terminator
20 chemistry to verify the sequences before overexpression.

21

22 *Protein expression and purification*

23 *E. coli* (BL21DE3) with the recombinant pETduet plasmid containing either native or
24 mutated PGM cDNA sequences were grown into a LB medium supplied with 100 μ g/mL
25 ampicillin at 37°C until they reach an absorbance of 0.6 at 600 nm. Protein expression was

1 induced by adding 1mM IPTG to the medium and kept at 37°C under shaking for 4 hours.
2 Cells were then harvested by centrifugation (4°C/15 000 g/5 min), and the pellets were re-
3 suspended in a binding buffer (20 mM Tris-HCl, pH 6.5, 500 mM NaCl, 5 mM imidazole),
4 disrupted by French Press at 1.6 kbar. After removing cell debris by centrifugation (15 000
5 g/4°C/60 min), supernatants (1 µg/mL of lysate) were treated with DNase I (Eurogentec) for
6 1 hour on ice. A first purification step was performed using immobilized metal affinity
7 chromatography with a His-bind resin column (His-Bond kit, Novagen) to recover PGM
8 variants, Protein binding with 5 mM and 60 mM imidazole and final elution of allozymes
9 with 1 M imidazole were performed following a classical chromatography protocol (pH 6.5).
10 The eluted fractions were concentrated using 30 kDa molecular cut-off Amicon-Ultra
11 (Millipore™). A second purification step was performed by size-exclusion chromatography
12 (SEC) with Superdex 75 column (1 x 30) (GE Healthcare) at a flow rate of 0.5 mL/min
13 monitored at 280 nm using a 25 mM Na₂HPO₄/NaH₂PO₄, pH 6.5. The purity of proteins was
14 checked by SDS-PAGE stained with Coomassie brilliant blue before being kept at 4°C in an
15 elution buffer supplemented with dithiothreitol (DTT, 10 mM) until use for enzyme assays.
16 The protein concentrations were measured by absorption at 280 nm with the theoretical
17 coefficient of 48,820 M⁻¹.cm⁻¹ as calculated using the ExPASy-ProtParam tool
18 (<http://web.expasy.org/protparam/>).

19

20 *Enzyme activity assay*

21 PGMs activities were assayed by coupling the formation of α-D-glucose 6-phosphate
22 (G6P) from α-D-glucose 1-phosphate (G1P) to NADPH formation using glucose 6-phosphate
23 dehydrogenase (G6PD) as a relay enzyme. The reaction mixture contained 50 mM Tris-HCl,
24 pH 7.4, 0.5 M MgCl₂, 1.2 mM NADP, 0.1 µM G6PD. The recombinant PGMs were used at
25 the following concentrations: [PGM₇₈] = 0.9 µM, [PGM₉₀] = 4 µM and [PGM₁₀₀] = 0.6 µM.

1 The concentration of the substrate (G1P) was varied from 0.2 to 60 mM to determine the
2 kinetic constants K_m and V_{max} using a Lineweaver-Burk plot.

3

4 *Thermal inactivation*

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

$$11 \quad y = \frac{(y_N + m_N \cdot T) + (y_D + m_D \cdot T) \cdot \exp\left(\frac{m(T - T_m)}{RT}\right)}{1 + \exp\left(\frac{m(T - T_m)}{RT}\right)} \quad [58] \quad (1)$$

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

17

18 *Guanidinium chloride-induced unfolding of PGM isoforms*

19 Unfolding of the PGM isoforms was induced by guanidinium chloride (GdmCl) in a 25 mM
20 sodium phosphate buffer, pH 6.5, NaCl 200 mM buffer. Proteins (12 μ M) were incubated
21 with increasing concentrations of GdmCl from 0 to 5 M, 30 min at 20°C and their intrinsic
22 fluorescence emission was determined at 324 nm under excitation at 290 nm on a Safas
23 Xenius spectrofluorimeter (Safas, Monaco). The GdmCl concentration was determined by
24 refractive index measurements^[59]. Biphasic states of protein denaturation with an intermediate

1 state (I) between native (N) and unfolded (U) states according to the following equilibrium:
2 $N \leftrightarrow I \leftrightarrow U$ were treated as follow: It was assumed that each transition ($N \leftrightarrow I$ and $I \leftrightarrow U$)
3 followed a two-state model of denaturation. The denatured protein fraction for each transition,
4 $f(I)$ for transition ($N \leftrightarrow I$) and $f(II)$ for transition ($I \leftrightarrow U$), was determined by resolving the two
5 following equations:

$$6 \quad f(I) = (y_N - y) / (y_N - y_I)$$

$$7 \quad f(II) = (y_I - y) / (y_I - y_U)$$

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

$$14 \quad f(I \text{ or } II) = \frac{\exp\left(-m \frac{(C_m - [GdmCl])}{RT}\right)}{1 + \exp\left(-m \frac{(C_m - [GdmCl])}{RT}\right)} \quad [60] \quad (2)$$

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

20

$$21 \quad \Delta G_{H2O}^{\circ} = m C_m \quad [58] \quad (3).$$

22

23 *3D PGM Structure Modelling*

1 PGM 78, 90 and 100 3D protein conformations were modelled with the Modeller
2 9v13^[61], using the structure of the crystallized rabbit phosphoglucomutase with its substrate
3 α -D-glucose 1-phosphate as a template (pdb file 1C47). This protein comprises 561 amino
4 acids with a resolution of 2.70 Å that shares 65% sequence identity with that of *Alvinella*
5 *pompejana*. One hundred models were generated for each PGM isoform and their quality was
6 assessed using the Modeller Objective Function parameter. Finally, a structure optimization
7 was obtained using the repair function of the FoldX software^[62].

8

9 RESULTS

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

11

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

25

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

3
4 In order to address the relationship between the two QE substitutions depicted from
5 the cDNA sequences and allozymes, direct sequencing (and/or RFLP) were performed on a
6 portion of exon 3 (94 codons) containing the double diagnostic mutations EQ in 220
7 individuals from both sides of the East Pacific Rise previously genotyped at the PGM-1
8 enzyme. The linkage disequilibrium between the two EQ mutations at codon positions 155
9 and 190, and allozymes was highly significant (Table 1) with correlation coefficients (R_{ij})
10 greater than 70% (p-values < 0.0001). This provides a very reliable correlation in which
11 combinations QE, EQ and EE correspond to the isoforms 78, 90 and 100, respectively. The
12 most negatively-charged allozyme 112, which is rare and always found at the heterozygous
13 state in the northern populations was also assigned to genotype EE, suggesting that an
14 additional replacement is occurring elsewhere in the protein. From this genotyping, groups of
15 individuals from either the North or the South EPR did not depart significantly from the
16 Hardy-Weinberg proportions. However, observed and expected heterozygosities were both
17 greater in the northern population ($H_{o-North}$: 0.40 vs $H_{o-South}$: 0.29). Interestingly, allele QQ
18 was not found in any of the populations. The frequencies of EE, EQ and QE alleles at the
19 sampled localities are summarized in Table S2. A more thorough analysis of the North/South
20 genetic differentiation was conducted on the 374 allelic sequences obtained by direct
21 sequencing (see alignment in supplementary data). The resulting haplotype network (Fig. S2)
22 shows a quasi-complete isolation of the northern and southern populations with a F_{st} value of
23 0.510 (see Table 2). Based on the 282 bp alignment, PGM90 (EQ) found in the Southern
24 population derives directly from the northern PGM90 (EQ) by one fixed mutation and the
25 southern PGM78 (QE) differs by 2 mutations from the northern PGM100 (EE). The

1 haplotype network also indicated that at least three alleles sampled in the southern
2 populations originated from the northern populations, suggesting that the barrier to gene flow
3 is not completely sealed.

4

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

6

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

16 Besides the two QE changes affecting the net charge of the protein in exon 3, other,
17 less common, cryptic amino-acid replacements were found along several regions of the *AP-*
18 *Pgm-1* CDS. This allowed us to estimate gene diversities and the south/north divergence over
19 an overall portion of about 3 kb (two thirds of the gene, see Table 2). Gene diversities were
20 almost constant over the *AP-Pgm-1* gene but allele divergence increases locally in the vicinity
21 of the two segregating EQ sites (Table 2). Looking more closely at the site variation along the
22 gene using a sliding window on our set of sequenced fragments indicated that gene diversity
23 is also slightly higher in exon3 where the two QE substitutions are found with values almost
24 identical to those depicted in intron2 (Fig. 3). This slight increase corresponded to peaks of
25 positive Tajima's D values, which raised up to +0.5 at the beginning of exon3, suggesting that

1 the presence of the two linked non-synonymous mutations may be associated with a hotspot
2 of gene diversity. Observed genetic diversities as estimated from θ_W and π were however not
3 significantly greater than expected from neutral coalescent simulations for both the southern
4 and northern populations over all the investigated *Pgm1* fragments (Fig.3, Table 2). Together
5 with the two QE variant sites, the genotyping of exon 1 also confirmed the occurrence of a
6 trans-equatorial V40L substitution found at a frequency of 0.15 restricted to the southern EQ
7 allele (PGM90) and one of the two northern allelic lineages, irrespective of the mutations EE
8 (PGM100) and EQ (PGM90). The direct sequencing of the two other genic regions located
9 either between exons 4 and 5 and between exons 6 and 8 did not show any additional
10 diagnostic amino-acid changes between the 3 allelic lineages EE, EQ and QE. By contrast,
11 several synonymous changes and indels appear to segregate between different allelic lineages
12 along the gene (see sequence alignments provided as supplementary data for exons 1, 3, 4, 5,
13 7 and introns 2, 6, and 7).

14

15 Estimating allele divergence and linkage disequilibria between segregating sites

16

17 To examine more specifically allele divergence and linkage disequilibria between
18 segregating sites within allelic lineages, a sequencing of recaptured alleles was targeted on the
19 longest region of the *AP-Pgm-1* gene (1110 bp). This region containing intron 2 and the two
20 allozyme-diagnostic sites EQ on exon 3 was thus genotyped from a subset of individuals. The
21 sequencing of 48 alleles highlighted the presence of a high level of synonymous
22 polymorphism with a strong linkage disequilibrium between these sites (Table 2), and two
23 diagnostic indels in intron 2 (insertions referred to as A and B following their order in the
24 intron). These segregating sites and indels allowed us to determine 4 divergent allelic lineages
25 with a few recombinants between them. These alleles were split between the northern and

1 southern populations. In the southern population, the two allelic lineages L1 and L2 refer to
2 the allozyme-diagnostic double mutation QE and EQ, whereas allelic lineages L3 and L4 refer
3 to EQ and a mixture of EQ and EE in the northern population (Fig. 4). At least, 9 and 11
4 synonymous mutations were fixed in intron 2 between allelic lineages L1 and L2, on one
5 hand, and L3 and L4, in the other hand.

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

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

17 To estimate the recombination rate, we examined the distribution of the *Rho* parameter
18 $4N_r$ with Phase 2.1.1 over a greater proportion of the gene (1-2860 bp) using segregating
19 sites (n=53) shared between northern and southern individuals that were successfully
20 sequenced for all exon-intron fragments of the *AP-Pgm-1* gene (N=20). Results from the
21 Phase -recomb and -hotspot outputs clearly indicated that the recombination rate between
22 alleles remains extremely low all along the gene (average local *Rho*= 0.033 and 0.008 for the
23 northern and southern populations, respectively, which further increased to nearly 2 in the
24 southern population at the end of the gene near the position 2140. This study therefore
25 indicated that the 4 allelic lineages greatly diverge one to each other in the vicinity of the

1 double mutation characterizing allozymes, with divergence even greater between allelic
2 sequences of the same population (0.7-1%) than those of the two geographic regions
3 investigated (0.9%).

4

5 To test whether the *AP-Pgm1* genetic patterns may be maintained by selection,
6 population parameters of both the northern and southern populations were simulated using a
7 msms structured coalescent with and without selection. Simulations indicated that a low
8 asymmetrical migration across the equatorial barrier with low or no recombination does not
9 explain by itself the observed patterns of genetic diversities found for the *AP-Pgm-1* gene
10 (Table S3). Simulated F_{st} values were around 0.8 and asymmetrical deme diversities were at
11 least two times smaller than the observed ones ($\pi = \theta_w = 3$) with and without recombination.
12 In this context, Tajima's D was close to zero within each deme as observed but highly
13 positive (+2.75) for the overall population when the observed one was also close to zero.
14 Introducing selection led to a better fit of simulated parameters to the observed ones.
15 Simulations with overdominance and low recombination led to a slight decrease of F_{st} values
16 (0.7) between demes, an increase of the within-deme genetic diversities close to the observed
17 ones but also produced greater positive Tajima's D (+1.3 for each deme). Our best fit to
18 values observed in the worm's populations was obtained with the two-niches model
19 simulations ($F_{st}=0.45$, converging nucleotidic diversities ($\theta_w=17 \rightarrow 13$) and Tajima'D (+0.8-
20 $>+0.4$) estimates within and between demes). These simulated values were even closest to
21 those estimated in the vicinity of the two EQ sites (intron2, see Tables 2, S3).

22

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

24

1 The obtention of full-length *AP-Pgm-1* cDNAs allowed us to examine the direct effect
2 of the two QE substitutions on the thermal stability and efficiency of the PGM-1 enzyme
3 using *in vitro* directed mutagenesis. To determine the conformational stability of the three
4 recombinant isoforms of the PGM-1, their guanidinium chloride (GdmCl)-induced unfolding
5 was studied. Variations of fluorescence with increasing concentrations of GdmCl were
6 biphasic (Fig S3) suggesting that the protein follows a three-state model of denaturation. For
7 each transition, the unfolded fraction of protein (f_u) was determined (Fig S4) and Gibbs free
8 energy change associated with each transition calculated (Table 3). For the two transitions,
9 the PGM90 (EQ) appears more stable than the two other isoforms. PGM78 (QE) appears
10 more stable than PGM100 for the first transition ($\Delta G^\circ_{H2O} = 8.0 \pm 0.46 \text{ kJ.mol}^{-1}$ vs. $\Delta G^\circ_{H2O} =$
11 $6.06 \pm 0.81 \text{ kJ.mol}^{-1}$ respectively), but not for the second transition ($\Delta G^\circ_{H2O} = 15.43 \pm 0.93$
12 kJ.mol^{-1} vs. $\Delta G^\circ_{H2O} = 15.13 \pm 0.98 \text{ kJ.mol}^{-1}$ respectively). The T_m values obtained from the
13 theoretical curve fitted on the thermal inactivation experimental data (inset Fig. 5) are very
14 similar for PGM78 ($46.5 \pm 1.7^\circ\text{C}$) and PGM100 ($44.0 \pm 0.1^\circ\text{C}$), but markedly higher for PGM90
15 ($50.9 \pm 0.7^\circ\text{C}$).

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

22

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

25

1 The Pompeii worm females exhibited an average coelomic fecundity of 200,000
2 oocytes with a great variability among them (values ranged from 1200 to 450 000 oocytes
3 depending on size (age) and the reproductive state^[41]). As opposed to our expectations,
4 females carrying the allele 90 were on average more fecund than homozygous females
5 carrying alleles 78 and 100. However, distributions of fecundity corrected by the size of the
6 female were not significantly different one to each other according to *Pgm-I* genotypes (One-
7 way ANOVA: $F=1.08$, $p=0.37$), see Fig. S5). This finding clearly indicates that the ability to
8 live under hotter conditions is not counter-balanced by a lesser reproductive success, at least
9 for the females.

10

11 DISCUSSION

12

13 Based on allozyme data, Piccino *et al.*^[12] previously proposed that the enzyme
14 polymorphism of the Pompeii worm *Alvinella pompejana* may be balanced at the locus *Pgm-*
15 *I*, at least in populations of the northern EPR. They indeed showed that allozymes 90 and 100
16 display distinct thermal stabilities and kinetic optima, the frequency of the most thermostable
17 isoform (allozyme 90) being positively correlated with temperature in newly-formed
18 edifices^[12,36]. Because the Pompeii worm is the only vent species able to colonize newly-
19 formed still ‘hot’ hydrothermal chimneys, bearing thermostable alleles is likely to represent
20 an adaptive advantage. The maintenance of polymorphism by selection on thermostable
21 alleles is however a matter of questioning. If advantageous in the hottest part of the vent
22 environment, thermostable alleles are indeed expected to rapidly spread in the population via
23 recurrent selective sweeps. This is obviously not the case for the *Pgm-I* locus, which
24 exhibited three major isoforms of different thermal stabilities (allozymes 78, 90 and 100) that
25 sharply segregate across a barrier to gene flow depicted by Plouviez *et al.*^[44] at the Equator.

1 Several hypotheses have thus been proposed by Piccino *et al.*^[12] about the maintenance of
2 alleles at the PGM-1 enzyme. This includes (i) allele overdominance due to the rapid
3 alternation of aerobic/anaerobic vent conditions, (ii) a fitness cost for individuals carrying the
4 most thermostable allele at this locus, or (iii) a two-niches model effect due to fluctuating
5 proportions of ‘hot’ and ‘cold’ habitats along the EPR.

6 In the present study, we sequenced the three major *Pgm-1* alleles of the worm to
7 investigate the distribution of non-synonymous polymorphisms along the gene, examine their
8 relationship with allozymes, and assess their evolutionary fate. We demonstrate that only two
9 linked mutations (E₁₅₅Q and E₁₉₀Q) are associated with the net charge of allozymes and also
10 responsible for the thermal performance of the three allozymes (78, 90 and 100). Looking at
11 the evolutionary history of these alleles indicates they are part of a rather ‘old’ polymorphism
12 that predates the vicariant event, which separated the EPR vent fauna across the Equator
13 about 1.5 Mya^[44, 65, 66]. In the following discussion, we therefore, examine arguments towards
14 the adaptive maintenance of the PGM-1 isoforms, and proposed that thermal compensation
15 represents a powerful mechanism by which different enzymatic properties might be
16 maintained under balancing selection - at least, during the exploration of the mutational
17 landscape of the protein that will lead to the emergence of the ‘optimal’ isoform - to optimize
18 metabolic fluxes as previously stated by Eanes^[67].

19

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

21 The non-synonymous polymorphism associated with the 4 allelic lineages of the *AP*
22 *Pgm-1* appears to be quite low ($\pi_N=0.0025$ on average). Such result sharply contrasts with
23 earlier studies on branch-point glycolytic enzymes that control the metabolic flux for
24 transport, storage and breakdown of carbohydrates, for which numerous cryptic non-
25 synonymous changes have been described^[67]. High levels of gene diversity were indeed

1 observed between the slow, medium and fast electrophoretic *Pgm-1* alleles of *Drosophila*
2 *melanogaster*^[14,68], or between phosphoglucose isomerase (PGI) alleles of *Colias*
3 butterflies^[69] suspected to evolve under balancing selection. By contrast, only eight non-
4 synonymous mutations (E37Q, V40L, E155Q, E190Q, R343I, G358S, T366M and F502L),
5 some of which are at a low frequency, have been detected from the direct comparison of
6 allelic sequences of *A. pompejana*. Moreover, even if gene diversity was slightly higher in the
7 intronic region preceding exon3 and in the exonic region where the two EQ sites are found, its
8 variation along the gene does not fit perfectly with the expectations of long-term balancing
9 selection. A weak ‘hot spot’ signal of silent site variation supported by slightly positive
10 Tajima’D and Fu&Li’F statistics is however observed near the doubly selected sites
11 E¹⁵⁵Q/E¹⁹⁰Q but not as strong as signals depicted for the *Adh* locus in *Drosophila*^[70,71].
12 Theoretical effects of balancing selection on nearby genome regions indeed promotes the
13 increase of genetic diversity near the selected site due to the lack of recombination and the
14 long-term accumulation of mutations^[38]. The very low level of nucleotidic polymorphism
15 found at the AP-*Pgm-1* can be however partially explained by recurrent population
16 bottlenecks due to the challenging environmental conditions that affect the whole vent
17 fauna^[15]. The joint action of abrupt demographic changes and habitat specialization should
18 indeed promote enzyme monomorphism. Under such conditions, the level of polymorphism
19 observed at the *Pgm-1*, although low, appears to be quite unusual when compared to most of
20 the genes examined in *A. pompejana*. In alvinellid worms, and especially thermophilic
21 species, proteins are indeed under strong purifying selection with overall d_N/d_S transcriptome
22 means very close to zero with values ranged between 0.02 and 0.05^[72]. To this extent, it is
23 worth noting that gene diversity at the *Pgm-1* locus appears to be locally two- to four-fold
24 higher than that recorded over the genome (ddRAD overall $\pi=0.0025$ ^[73]) and from other
25 reported genes^[35].

1 Looking more specifically at the 4 allelic lineages of the *AP-PgmI* nearby the EQ sites
2 (intron 2) clearly indicates that they have accumulated a great number of synonymous
3 substitutions since their separation with almost no recombination events (low values of R_m and
4 Rho , see the Phase 2.1.1 analysis). The two allelic lineages L1 and L2 present in the Southern
5 population exhibit 1% divergence between them, with a strong linkage disequilibrium
6 between the two variant sites $E^{155}Q^{190}$ (PGM90) and $Q^{155}E^{190}$ (PGM78) and the silent
7 substitutions found in intron2. This suggests that these two allelic lineages evolved separately
8 without recombination in the vicinity of the two non-synonymous sites for a long period of
9 time. The two northern allelic lineages (L3 and L4) also diverged by 0.7% divergence and
10 two diagnostic indels in intron2. These silent mutations and indels are also linked together,
11 suggesting once more that these two lineages evolved separately but for a shorter period of
12 time. However, diagnostic mutations are not completely linked to the variant sites $E^{155}E^{190}$
13 (PGM100) and $E^{155}Q^{190}$ (PGM90). Although L3 forms a single clade associated with $E^{155}Q^{190}$,
14 L4 is a mixture of $E^{155}E^{190}$ and $E^{155}Q^{190}$ suggesting either that, at least one recombination
15 event occurred between the two northern alleles or that $E^{155}E^{190}$ (PGM100) is a recently
16 derived variant in the northern population. Finally, both divergences observed between alleles
17 within each population are of the same amplitude as the divergence estimated between the
18 southern and northern alleles (0.9%), which coincide with an overall significant F_{st} value of
19 0.588 between them. If we accept that the southern (L1 and L2) and northern (L3 and L4)
20 allelic lineages become isolated after the appearance of the physical barrier to dispersal, about
21 1.2 Mya^[35,44], this clearly indicates that the co-occurrence of the 4 highly divergent *Pgm-I*
22 alleles derives from an older polymorphism predating the vicariant event that separated the
23 Northern and Southern vent fauna of the East Pacific Rise, with the possible emergence of the
24 genotype $E^{155}E^{190}$ in the North. Such a scenario is likely confirmed by the distribution of
25 alleles in the exon3 haplotype network and the signature of the linked silent polymorphic sites

1 in introns 1 and 2, where northern alleles seem to derive from a southern allele bearing the
2 Q¹⁵⁵E¹⁹⁰ 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 about the selective modalities acting on the co-occurrence of alleles when one of the
8 two alleles is better adapted to high temperatures^[12]. One of the first explanations to the
9 adaptive maintenance of AP-PGM1 isoforms was to consider that the rapid alternation of oxic
10 and anoxic conditions during venting should favor heterozygote excesses if the heterozygote's
11 fitness is close to that of the favored homozygote in one of the two habitats^[74]. Pogson^[13]
12 previously proposed that overdominance represents the most likely evolutionary mechanism
13 at the origin of the maintenance of a balanced polymorphism at the *PGM-2* locus for the
14 oyster *Crassostrea gigas*, (but also see^[75] for its link with individual's growth rate). Here, we
15 were not able to detect any overdominance at the *Pgm-1* locus in terms of heterozygote
16 excesses in natural populations, and simulations of structured coalescent with asymmetrical
17 migration and overdominance always led to very high within-deme positive Tajima's D and
18 unequal diversities between demes not observed here. This finding confirms the previous
19 study^[12] with allozymes. It is however worth noting that such advantage can be easily masked
20 by the temporal dynamics of the thermal habitat (chimneys refreshing with time) and the
21 juxtaposition of chimneys of different ages. The worm is indeed exposed to a mosaic of
22 fluctuating thermal habitats where temperature could spatially vary according to the age of the
23 chimneys^[12].

24 Maintenance of allozymes with different thermal stabilities can be also explained by a
25 two-niches model of local differentiation with habitat and drift^[76,77,78,79]. Simulating

1 coalescences with a 2-niches model with a theta value equal to the observed value indeed
2 provided parameters estimates (Fst, overall and intra-deme diversities with Tajima's D) much
3 closer to values observed in the vicinity of the two EQ sites than the two other models of
4 asymmetric gene flow with and without overdominance. Given the spatial and temporal
5 dynamics of the hydrothermal discharge, changes in the frequency of *Pgm-1* alleles could be
6 either due to local adaptation or an exacerbated genetic drift associated with the dynamics of
7 colonization of the newly opened sites. Indeed, the dynamic nature of hydrothermal vents
8 over longer time scales (years) led to a very patchy and transient habitat, scattered along the
9 EPR with a complex heterogeneity of age-driven vent conditions. This can be seen as a
10 multitude of distinct ecological niches for the same species. In this context, the proportion of
11 newly formed 'still hot' chimneys and older colder ones greatly varies over time depending
12 on the spreading rate of the rift and thus the frequency of tectonic and volcanic events along
13 the East Pacific Rise.

14 Piccino *et al.*^[12] also proposed that the maintenance of a bi-allelic polymorphism at the
15 *Pgm-1* locus results from a fitness cost to the colonization of the early stages of a chimney.
16 The first settlers on 'hot' (>100°C) anhydrite chimneys indeed benefit from a lack of
17 predators and competitors. Colonists may therefore display more thermoresistant alleles but a
18 lesser reproductive investment and/or survival. Watt and collaborators^[80,81,82] and Wheat *et*
19 *al.*^[69] previously showed that both PGI and PGM have a great contribution on the male mating
20 fitness in *Colias* butterflies, probably as the result of longer and more vigorous flight within
21 the day. Based on female fecundity, we were however not able to observe fitness differences
22 between the *Pgm-1* genotypes. This suggests that a better predisposition to colonize still 'hot'
23 chimneys is probably not compensated by a reduced reproductive success to prevent the
24 fixation of the advantageous allele. In the fruitfly, the *Pgm* locus represents however a
25 quantitative trait for glycogen storage and hence, the ability to survive better to starvation^[34].

1 In the case of *A. pompejana*, differences in the thermal regime could be great between
2 colonists and reproducers. As a consequence, colonists subjected to longer periods of high
3 temperature (and associated hypoxia) may be maladapted to produce and use glycogen
4 reserves and, thus, may not invest much into reproduction. On the contrary, secondary settlers
5 arriving in much cooler conditions are more likely to use their glycogen reserves to massively
6 invest in the production of gametes as previously shown by Faure *et al.*^[41]. Colder conditions
7 indeed seem to be a prerequisite for releasing fertilized eggs after pairing as embryos are not
8 able to develop at temperatures greater than 15°C^[40].

9

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

11 In *Drosophila*, *Pgm-1* variants play a non-negligible role in the regulation of the
12 metabolic energy pool along latitudinal clines where the decrease of temperature is
13 compensated by an increase of the enzyme activity. Populations of *Drosophila melanogaster*
14 living at the highest (and thus coldest) latitudes possess PGM allozymes with a higher
15 catalytic efficiency and greater glycogen contents. According to the theory of metabolic flux,
16 this can be an adaptive way of temperature compensation to maintain the same glycogen
17 contents over the latitudinal gradient^[67]. Differences in both protein thermostability and
18 catalytic efficiency between *Pgm* alleles were previously reported to explain both local
19 differentiation and latitudinal clines in the oyster *Crassostrea gigas*^[6,13] and *Drosophila*
20 *melanogaster*^[34]. By comparison, thermal compensation may be therefore directly linked to
21 different biochemical phenotypes that interact with the growth rate and reproductive effort of
22 the worms. The theory predicts that differences in activity at only one enzyme must be
23 however substantial to affect metabolic fluxes between genotypes^[83]. To test this hypothesis,
24 one could measure the effect of non-synonymous mutations on the functional properties of the
25 enzyme, its conformational stability and their effect on population fitness. In this study, three

1 recombinant isoforms of PGM-1 ($E^{155}E^{190}$ (PGM100), $E^{155}Q^{190}$ (PGM90) and $Q^{155}E^{190}$
2 (PGM78)) were obtained by directed mutagenesis. The replacement of the glutamate by a
3 glutamine at position 190 increases the conformational stability and thermostability of the
4 protein, confirming that PGM90 is the most thermostable isoform. As discussed by Piccino *et*
5 *al.*^[12], carrying this allele may be advantageous during the colonization of newly-formed
6 chimneys whose surface temperature usually exceeds 50°C. Unexpectedly, PGM90 exhibits a
7 decrease in the catalytic efficiency of the enzyme when compared to the two other
8 recombined variants, (k_{cat}/K_m is a hundred times for PGM78 and nearly two-fold greater for
9 PGM100 when compared to PGM90). The recombinant PGM90 also exhibits the lowest
10 affinity for its substrate glucose-1-phosphate at 17°C. This finding is of importance because
11 such a genetically-determined trade-off between protein stability and enzyme activity was not
12 reported in other invertebrate species subjected to balancing selection so far^[6,13,34,82].
13 Increased thermostability of a protein is often associated with a decrease in the flexibility of
14 the molecule, and thus the dynamics of the enzyme reaction^[84,85]. Our results are in perfect
15 agreement with these theoretical expectations and support the positive role of a
16 thermodynamic trade-off between thermostability and catalysis as previously proposed by
17 Eanes^[11] to explain the co-occurrence of alleles. PGM90 can remain stable for a longer period
18 of time but is less efficient to either produce or consume the glycogen reserves of the worm
19 than the two other isoforms are. To this extent, the fact that the k_{cat}/K_m ratio of isoform
20 PGM78 is much higher than that of the isoform PGM100 can explain why isoform 78 is more
21 frequent in the Southern populations (about 80%) when compared with isoform PGM100 in
22 the Northern populations (around 70%). The balance between allele frequencies from both
23 sides of the Equator may be dictated by the selective coefficient attributed to each genotype as
24 the direct reflection of the catalytic efficiency difference between PGM90 and its alternative
25 isoforms.

1

2 *Structural effect of mutations E155Q and E190Q*

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

1 to greater sensitivity to temperature. Finally, the glutamine replacement at position 155 is also
2 likely to play a key role in the molecular dynamics of the protein, especially during the 180°
3 rotation of the reaction intermediate (glucose-1,6-diphosphate) inside the active site. This
4 likely explains the higher enzymatic efficiency of the isoform 78.

5

6 Acknowledgments

7 This article has been written in memory of Dominique Le Guen, who greatly helped us in
8 setting up of the allozyme genotyping. We thank chief scientists and the ‘Nautile’ crews for
9 their technical support and efforts during the oceanographic expeditions Phare2002,
10 Biospedo2004 and Mescal2010. We are very grateful to the two anonymous referees who
11 provided valuable comments and editorial suggestions on the manuscript.

12

13 Literature Cited

- 14 [1] Holt, R. D., and M. S. Gaines, 1992 Analysis of adaptation in heterogeneous landscapes:
15 implications for the evolution of fundamental niches. *Evolut. Ecol.* 6: 433-447.
- 16 [2] Schmidt, P. S., E. A. Serrão, G. A. Pearson, C. Riginos, P. D. Rawson, *et al.*, 2008
17 Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific
18 loci. *Ecology* 89: 91-107.
- 19 [3] Nevo, E., T. Shimony, and M. Libni, 1977 Thermal selection of allozyme polymorphisms
20 in barnacles. *Nature*, 267(5613): 699.
- 21 [4] Nevo, E., E. Lavie, and R. Ben-Shlomo, 1983 Selection of allelic isozyme polymorphisms
22 in marine organisms: pattern, theory, and application. *Isozymes*, 10: 69-92.
- 23 [5] Nevo, E., R. Noy, B. Lavie, A. Beiles, and S. Muchtar, 1986 Genetic diversity and
24 resistance to marine pollution. *Biol. J. Linn. Soc.* 29: 139-144.

- 1 [6] Pogson, G. H., 1989 Biochemical characterization of genotypes at the
2 phosphoglucosyltransferase-2 locus in the Pacific oyster, *Crassostrea gigas*. *Biochem. Genetics*, 27:
3 571-589.
- 4 [7] Riddoch, B., 1993 The adaptive significance of electrophoretic mobility in
5 phosphoglucose isomerase (PGI). *Biol. J. Linn. Soc.* 50: 1-17.
- 6 [8] Schmidt, P. S., and D. M. Rand, 2001 Adaptive maintenance of genetic polymorphism in
7 an intertidal barnacle: habitat-and life-stage-specific survivorship of MPI genotypes.
8 *Evolution* 55: 1336-1344.
- 9 [9] Boutet, I., A. Tanguy, D. Le Guen, P. Piccino, S. Hourdez, *et al.*, 2009 Global depression
10 in gene expression as a response to rapid thermal changes in vent mussels. *Proc. Roy. Soc.*
11 *London B* 276: 3071-3079.
- 12 [10] Bougerol, M., I. Boutet, D. Le Guen, D. Jollivet, and A. Tanguy, 2015 Transcriptomic
13 response of the hydrothermal mussel *Bathymodiulus azoricus* mediated by heavy metals is
14 modulated by *Pgm* genotypes and symbiont content. *Mar. Genomics*,
15 <http://dx.doi.org/10.1016/j.margen.2014.11.010>.
- 16 [11] Eanes, W. F., 1999 Analysis of selection on enzyme polymorphisms. *Ann. Rev. Ecol.*
17 *Syst.* 30: 301-326.
- 18 [12] Piccino, P., F. Viard, P.-M. Sarradin, N. Le Bris, D. Le Guen, *et al.*, 2004 Thermal
19 selection of PGM allozymes in newly founded populations of the thermotolerant vent
20 polychaete *Alvinella pompejana*. *Proc. Roy. Soc. London B* 271: 2351-2359.
- 21 [13] Pogson, G. H., 1991 Expression of overdominance for specific activity at the
22 phosphoglucosyltransferase-2 locus in the Pacific oyster, *Crassostrea gigas*. *Genetics*, 128: 133-141.
- 23 [14] Verrelli, B. C., and W. F. Eanes, 2001a Clinal variation for amino acid polymorphisms at
24 the *Pgm* locus in *Drosophila melanogaster*. *Genetics* 157: 1649-1663.

- 1 [15] Vrijenhoek, R. C., 1997 Gene flow and genetic diversity in naturally fragmented
2 metapopulations of deep-sea hydrothermal vent animals. *J. Hered.* 88: 285-293.
- 3 [16] Jollivet, D., P. Chevaldonné, B. Planque, 1999 Hydrothermal-vent alvinellid polychaete
4 dispersal in the Eastern Pacific. 2. A metapopulation model based on habitat shifts. *Evolution*
5 53: 1128-1142.
- 6 [17] Vrijenhoek, R. C., 2010 Genetic diversity and connectivity of deep-sea hydrothermal
7 vent metapopulations. *Mol. Ecol.* 19: 4391-4411.
- 8 [18] Watremez, P., and C. Kervevan, 1990 Origine des variations de l'activité hydrothermale:
9 premiers éléments de réponse d'un modèle numérique simple. *C. R. Acad. Sci Paris, Sér. 2,*
10 311: 153-158.
- 11 [19] Johnson, K. S., J. J. Childress, R. R. Hessler, C. M. Sakamoto-Arnold, and C. L. Beehler,
12 1988 Chemical and biological interactions in the Rose Garden hydrothermal vent field,
13 Galapagos spreading center. *Deep Sea Res. A* 35: 1723-1744.
- 14 [20] Le Bris, N., and F. Gaill, 2007 How does the annelid *Alvinella pompejana* deal with an
15 extreme hydrothermal environment? *Rev. Environ. Sci. Biotechnol.* 6: 197-221.
- 16 [21] Hourdez, S., and Lallier, F. H. 2006 Adaptations to hypoxia in hydrothermal-vent and
17 cold-seep invertebrates. In *Life in extreme environments*. Springer, Dordrecht, pp. 297-313.
- 18 [22] Von Damm, K. L., 1995 Controls on the chemistry and temporal variability of seafloor
19 hydrothermal fluids. pp. 222-247 in *Seafloor hydrothermal systems: Physical, chemical,*
20 *biological, and geological interactions*, 91.
- 21 [23] Jollivet, D., 1996 Specific and genetic diversity at deep-sea hydrothermal vents: an
22 overview. *Biodiv. Cons.* 5: 1619-1653.
- 23 [24] Matabos, M., N. Le Bris, S. Pendlebury, and E. Thiébaud, 2008 Role of physico-chemical
24 environment on gastropod assemblages at hydrothermal vents on the East Pacific Rise (13
25 N/EPR). *J. 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
2 Temporal and spatial patterns of biological community development at nascent deep-sea
3 hydrothermal vents (9 50' N, East Pacific Rise). *Deep Sea Res. II: Topical Studies in*
4 *Oceanography* 45: 465-515.
- 5 [26] Fontaine, F. J., M. Cannat, and J. Escartin, 2008 Hydrothermal circulation at slow-
6 spreading mid-ocean ridges: The role of along-axis variations in axial lithospheric thickness.
7 *Geology* 36: 759-762.
- 8 [27] Marcus, J., V. Tunnicliffe, and D. A. Butterfield, 2009 Post-eruption succession of
9 macrofaunal communities at diffuse flow hydrothermal vents on Axial Volcano, Juan de Fuca
10 Ridge, Northeast Pacific. *Deep Sea Res. II: Topical Studies in Oceanography*, 56: 1586-1598.
- 11 [28] Chevaldonné, P., D. Desbruyères, and J. J. Childress, 1992 ... and some even hotter.
12 *Nature* 359(6396): 593.
- 13 [29] Cary, S. C., T. Shank, and J. Stein, 1998 Worms bask in extreme temperatures. *Nature*
14 391(6667): 545.
- 15 [30] Ravaux, J., G. Hamel, M. Zbinden, A. A. Tasiemski, I. Boutet, *et al.*, 2013 Thermal limit
16 for metazoan life in question: *in vivo* heat tolerance of the Pompeii worm. *PLoS One*, 8(5):
17 e64074.
- 18 [31] Jang, S. J., E. Park, W. K. Lee, S. B. Johnson, R. C. Vrijenhoek, *et al.*, 2016 Population
19 subdivision of hydrothermal vent polychaete *Alvinella pompejana* across equatorial and
20 Easter Microplate boundaries. *BMC Evolut. Biol.* 16: 235.
- 21 [32] Desbruyères, D., P. Chevaldonné, A.-M. Alayse, D. Jollivet, F. H. Lallier, *et al.*, 1998
22 Biology and ecology of the “Pompeii worm”(*Alvinella pompejana* Desbruyères and Laubier),
23 a normal dweller of an extreme deep-sea environment: a synthesis of current knowledge and
24 recent developments. *Deep Sea Res. II: Topical Studies in Oceanography* 45: 383-422.

- 1 [33] Chevaldonné, P., C. R. Fisher, J. J. Childress, D. Desbruyères, D. Jollivet, *et al.*, 2000
2 Thermotolerance and the ‘Pompeii worms’. *Mar. Ecol. Prog. Ser.* 208: 293-295.
- 3 [34] Jollivet, D., D. Desbruyères, F. Bonhomme, and D. Moraga, 1995a Genetic
4 differentiation of deep-sea hydrothermal vent alvinellid populations (Annelida: Polychaeta)
5 along the East Pacific Rise. *Heredity*, 74: 376-391.
- 6 [35] Plouviez, S., D. Le Guen, O. Lecompte, F. H. Lallier, and D. Jollivet, 2010 Determining
7 gene flow and the influence of selection across the equatorial barrier of the East Pacific Rise
8 in the tube-dwelling polychaete *Alvinella pompejana*. *BMC Evol. Biol.*, 10: 220.
- 9 [36] Jollivet, D., D. Desbruyères, C. Ladrat, and L. Laubier, 1995b Evidence for differences
10 in the allozyme thermostability of deep-sea hydrothermal vent polychaetes (Alvinellidae): a
11 possible selection by habitat. *Mar. Ecol. Prog. Ser.* 123:125-136.
- 12 [37] Verrelli, B. C., and W. F. Eanes, 2001b The functional impact of *Pgm* amino acid
13 polymorphism on glycogen content in *Drosophila melanogaster*. *Genetics* 159: 201-210.
- 14 [38] Charlesworth, D., 2006 Balancing selection and its effects on sequences in nearby
15 genome regions. *PLoS Genetics* 2: 379-384.
- 16 [39] Hedrick, P. W., 2007 Balancing selection. *Current Biol.* 17: 230-231.
- 17 [40] Pradillon, F., B. Shillito, C. M. Young, and F. Gaill, 2001 Deep-sea ecology:
18 Developmental arrest in vent worm embryos. *Nature*, 413(6857): 698.
- 19 [41] Faure, B., P. Chevaldonné, F. Pradillon, E. Thiébaud, E., and D. Jollivet, 2007 Spatial
20 and temporal dynamics of reproduction and settlement in the Pompeii worm *Alvinella*
21 *pompejana* (Polychaeta: Alvinellidae). *Mar. Ecol. Prog. Ser.* 348: 197-211.
- 22 [42] Mishra, R. N., S. L. Singla-Pareek, S. Nair, S. K. Sopory, and M. K. Reddy, 2002
23 Directional genome walking using PCR. *Biotechniques*, 33: 830-842.

- 1 [43] Bierne, N., A. Tanguy, M. Faure, B. Faure, E. David, E., *et al.*, 2007 Mark-recapture
2 cloning: a straightforward and cost-effective cloning method for population genetics of single
3 copy nuclear DNA sequences in diploids. *Mol. Ecol. Notes* 7: 562-566.
- 4 [44] Plouviez, S., T. M. Shank, B. Faure, C. Daguin Thiébaud, F. Viard, *et al.*, 2009
5 Comparative phylogeography among hydrothermal vent species along the East Pacific Rise
6 reveals vicariant processes and population expansion in the South. *Mol. Ecol.* 18: 3903-3917.
- 7 [45] Black, W. C., and E. S. Krafur, 1985 A FORTRAN program for the calculation and
8 analysis of two-locus linkage disequilibrium coefficients. *Theor. Appl. Genet.* 70: 491-496.
- 9 [46] Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Catch, 2004 *GENETIX 4.0. 5.2,*
10 *Software under Windows™ for the genetics of the populations.* University of Montpellier,
11 Montpellier, France.
- 12 [47] Bandelt, H. J., P. Forster, and A. Röhl, 1999 Median-joining networks for inferring
13 intraspecific phylogenies. *Mol. Biol. Evol.* 16: 37-48.
- 14 [48] Librado, P., and J. Rozas, 2009 DnaSP v5: a software for comprehensive analysis of
15 DNA polymorphism data. *Bioinformatics*, 25: 1451-1452.
- 16 [49] Kelly, J. K., 1997 A test of neutrality based on interlocus associations. *Genetics*, 146:
17 1197-1206.
- 18 [50] Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of
19 recombination events in the history of a sample of DNA sequences. *Genetics* 111: 147-164.
- 20 [51] Martin, D. P., P. Lemey, M. Lott, V. Moulton, D. Posada, *et al.*, 2010 RDP3: a flexible
21 and fast computer program for analyzing recombination. *Bioinformatics*, 26: 2462-2463.
- 22 [52] Li, N. and M. Stephens, 2003 Modelling linkage disequilibrium and identifying
23 recombination hotspots using snp data. *Genetics* 165, 2213-2233.
- 24 [53] Hudson, R. R., M. Slatkin, and W. P. Maddison, 1992 Estimation of levels of gene flow
25 from DNA sequence data. *Genetics* 132: 583-589.

- 1 [54] Kumar, S., G. Stecher, and K. Tamura, 2016 MEGA7: molecular evolutionary genetics
2 analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33: 1870-1874.
- 3 [55] Ewing, G., and J. Hermisson, 2010 MSMS: a coalescent simulation program including
4 recombination, demographic structure and selection at a single locus. *Bioinformatics*, 26(16):
5 2064-2065.
- 6 [56] Thornton K. 2003 Libsequence: a c++ class library for evolutionary genetic
7 analysis. *Bioinformatics*, 19(17) : 2325–2327.
- 8 [57] Reikofski, J., and B. Y. Tao, 1992 Polymerase chain reaction (PCR) techniques for site-
9 directed mutagenesis. *Biotechnol. Adv.* 10: 535-547.
- 10 [58] Pace, C. N., and J. M. Scholtz, 1997 Measuring the conformational stability of a protein.
11 *Protein structure: A practical approach*, 2: 299-321.
- 12 [59] Pace, C. N., 1986 Determination and analysis of urea and guanidine hydrochloride
13 denaturation curves. pp. 266-280 in *Methods in enzymology* Vol. 131, Academic Press.
- 14 [60] Walter, P., and D. Ron, 2011 The unfolded protein response: from stress pathway to
15 homeostatic regulation. *Science* 334: 1081-1086.
- 16 [61] Sali, A., and T. L. Blundell, 1993 Comparative protein modelling by satisfaction of
17 spatial restraints. *J. Mol. Biol.* 234: 779-815.
- 18 [62] Guerois, R., J. E. Nielsen, and L. Serrano, 2002 Predicting changes in the stability of
19 proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* 320: 369-
20 387.
- 21 [63] Whitehouse, D. B., W. Putt, J. U. Lovegrove, K. Morrison, M. Hollyoake, *et al.*, 1992
22 Phosphoglucomutase 1: complete human and rabbit mRNA sequences and direct mapping of
23 this highly polymorphic marker on human chromosome 1. *Proc. Nat. Acad. Sci. USA* 89:
24 411-415.

- 1 [64] Tanguy, A., I. Boutet, P. Boudry, L. Degremont, J. Laroche, et al., 2006 Molecular
2 identification and expression of the phosphoglucomutase (PGM) gene from the Pacific oyster
3 *Crassostrea gigas*. *Gene*, 382: 20-27.
- 4 [65] Matabos M., S. Plouviez, S. Hourdez, D. Desbruyères, P. Legendre, *et al.*, 2011 Faunal
5 changes and geographic crypticism indicate the occurrence of a biogeographic transition zone
6 along the Southern East-Pacific Rise. *J. Biogeogr.* 38: 575-594.
- 7 [66] Matabos, M., and D. Jollivet, 2019 Revisiting the species' complex of *Lepetodrilus*
8 *elevatus* (Vetigastropod, Lepetodrilidae) using gastropod samples from the Galápagos and
9 Guaymas hydrothermal vent systems. *J. Moll. Stud.* 85: 154-165.
- 10 [67] Eanes, W. F., 2011 Molecular population genetics and selection in the glycolytic
11 pathway. *J. Exp. Biol.* 214: 165-171.
- 12 [68] Verrelli, B. C., and W. F. Eanes, 2000 Extensive amino acid polymorphism at the *Pgm*
13 locus is consistent with adaptive protein evolution in *Drosophila melanogaster*. *Genetics*,
14 156: 1737-1752.
- 15 [69] Wheat, C. W., W. B. Watt, D. D. Pollock, and P. M. Schulte, 2005 From DNA to fitness
16 differences: sequences and structures of adaptive variants of *Colias* phosphoglucose
17 isomerase (PGI). *Mol. Biol. Evol.* 23: 499-512.
- 18 [70] McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus in
19 *Drosophila*. *Nature*, 351(6328), 652-654.
- 20 [71] Begun, D. J., Betancourt, A. J., Langley, C. H., & Stephan, W. (1999). Is the fast/slow
21 allozyme variation at the *Adh* locus of *Drosophila melanogaster* an ancient balanced
22 polymorphism?. *Molecular Biology and Evolution*, 16(12), 1816-1819.
- 23 [72] Fontanillas, E., O.V. Galzitskaya, O. Lecompte, M. Y. Lobanov, A. Tanguy, *et al.*, 2017
24 Proteome evolution of deep-sea hydrothermal vent alvinellid polychaetes supports the

- 1 ancestry of thermophily and subsequent adaptation to the cold for some lineages. *Genome*
2 *Biol. Evol.* 9: 279-296.
- 3 [73] Bioy, A., 2018 *Histoire évolutive et influence de la selection sur la diversité génétique*
4 *des annélides polychètes d'environnements extrêmes*. Thèse de Doctorat, Sorbonne
5 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
8 homozygote in each environment. *Genet. Res.* 45: 299-314.
- 9 [75] Gardner, J. P. A., and I. Lobkov, 2005 A test for overdominance at the
10 phosphoglucumutase-2 locus in Pacific oysters (*Crassostrea gigas*) from B-New Zealand.
11 *Aquaculture* 244: 29-39.
- 12 [76] Levene, H., 1953 Genetic equilibrium when more than one ecological niche is available.
13 *Am. Nat.* 87: 331-333.
- 14 [77] Gillespie, J. H., 1985 The interaction of genetic drift and mutation with selection in a
15 fluctuating environment. *Theor. Pop. Biol.* 27: 222-237.
- 16 [78] Hedrick, P. W., M. E. Ginevan, and E. P. Ewing, 1976 Genetic polymorphism in
17 heterogeneous environments. *Ann Rev Ecol. Syst.* 7: 1-32.
- 18 [79] Hedrick, P. W., 1986 Genetic polymorphism in heterogeneous environments: a decade
19 later. *Ann. Rev. Ecol. Syst.* 17: 535-566.
- 20 [80] Watt, W. B., 1977 Adaptation at specific loci. I. Natural selection on phosphoglucose
21 isomerase of *Colias* butterflies: biochemical and population aspects. *Genetics* 87: 177-194.
- 22 [81] Watt, W. B., R. C Cassin, and M. S. Swan, 1983 Adaptation at specific loci. III. Field
23 behavior and survivorship differences among *Colias* PGI genotypes are predictable from *in*
24 *vitro* biochemistry. *Genetics* 103: 725-739.

- 1 [82] Carter, P. A., and W. B. Watt, 1988 Adaptation at specific loci. V. Metabolically
2 adjacent enzyme loci may have very distinct experiences of selective pressures. *Genetics* 119:
3 913-924.
- 4 [83] Hartl, D. L., D. E. Dykhuizen, and A. M. Dean, 1985 Limits of adaptation: the evolution
5 of selective neutrality. *Genetics*, 111: 655-674.
- 6 [84] Somero, G. N., 1978 Temperature adaptation of enzymes: biological optimization
7 through structure-function compromises. *Ann. Rev. Ecol. Syst.* 9: 1-29.
- 8 [85] Somero, G. N., 1995 Proteins and temperature. *Ann. Rev. Physiol.* 57: 43-68.
- 9 [86] Vogt, G., S. Woell, P. Argos, 1997 Protein thermal stability, hydrogen bonds, and ion
10 pairs. *J. Mol. Biol.* 269: 631-643.
- 11 [87] Shin, D. S., M. DiDonato, D. P. Barondeau, G. L. Hura, C. Hitomi, *et al.*, 2009
12 Superoxide dismutase from the eukaryotic thermophile *Alvinella pompejana*: structures,
13 stability, mechanism, and insights into amyotrophic lateral sclerosis. *J. Mol. Biol.* 385: 1534-
14 1555.
- 15 [88] Bruneaux, M., J. Mary, M. Verheye, O. Lecompte, O. Poch, *et al.*, 2013 Detection and
16 characterisation of mutations responsible for allele-specific protein thermostabilities at the
17 Mn-superoxide dismutase gene in the deep-sea hydrothermal vent polychaete *Alvinella*
18 *pompejana*. *J. Mol. 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
 4 numbers of tests still significant after the Bonferonni correction, respectively. (RDP n.d.): recombinant not detected using automated RDP and bootscan
 5 packages 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
 6 coding 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
$Z_n S$	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}

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

1 Table 3. Conformational and temperature stability of the three overexpressed variants (PGM78, PGM90, and PGM100). C_m et m values
 2 estimated from the variation of protein fluorescence in presence of an increasing concentration of GdmHCl (values for each of the two
 3 transitions). Estimation of the free enthalpy of the unfolding reaction in absence of chaotropic agent for each of the two transition states. T_m :
 4 values of the temperature at which we reach 50% of non-reversible inactivation after a 30-minute exposure. K_m^{app} and K_{cat} are kinetic parameters
 5 corresponding to the apparent Michaelis-Menten constant for glucose-1-phosphate, and the catalytic constant, respectively. The ratio of these two
 6 values corresponds to the specific activity.
 7
 8

	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
$\Delta G_{H_2O}^0$ (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+} .

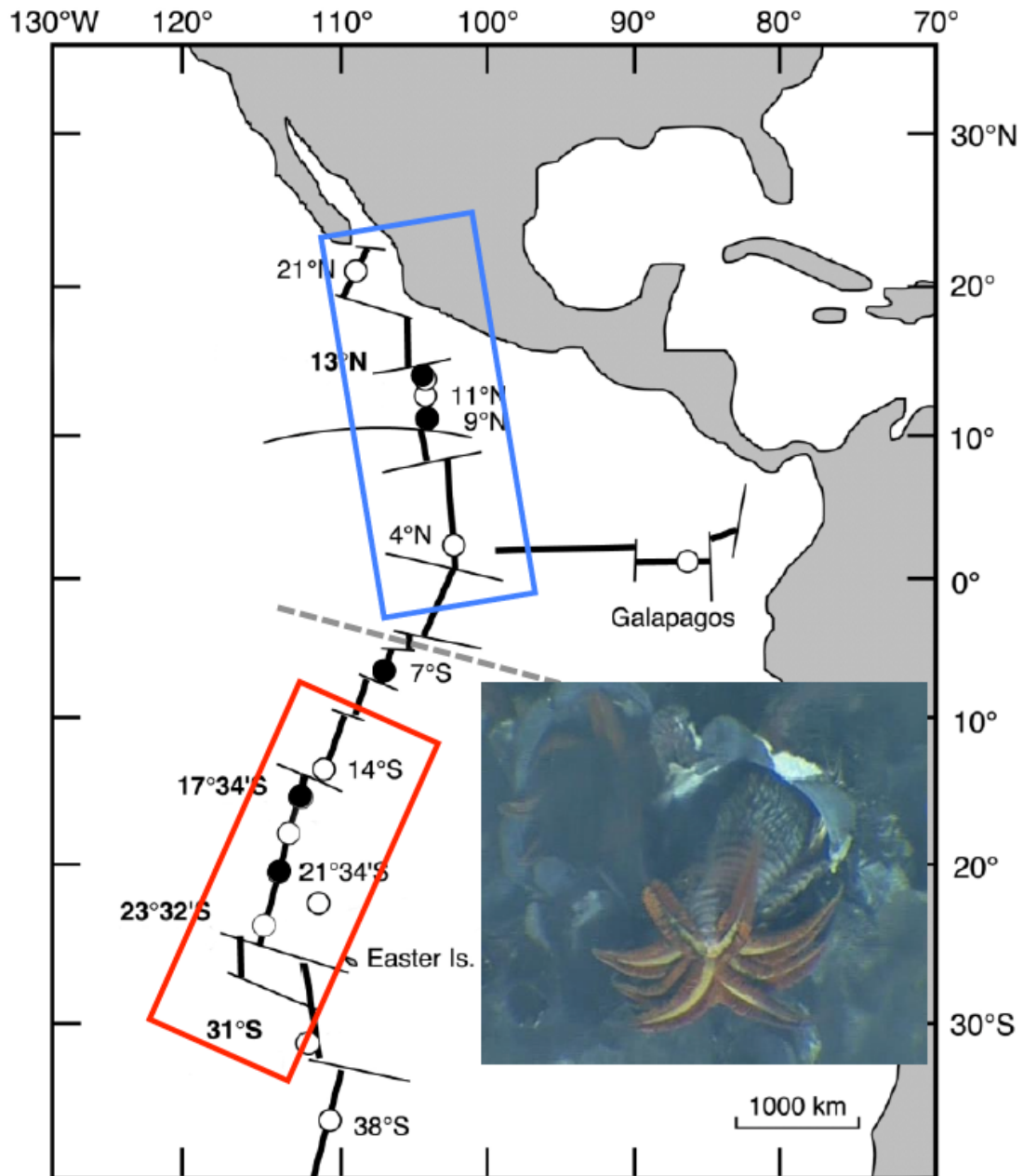


Fig. 1

Fig. 2

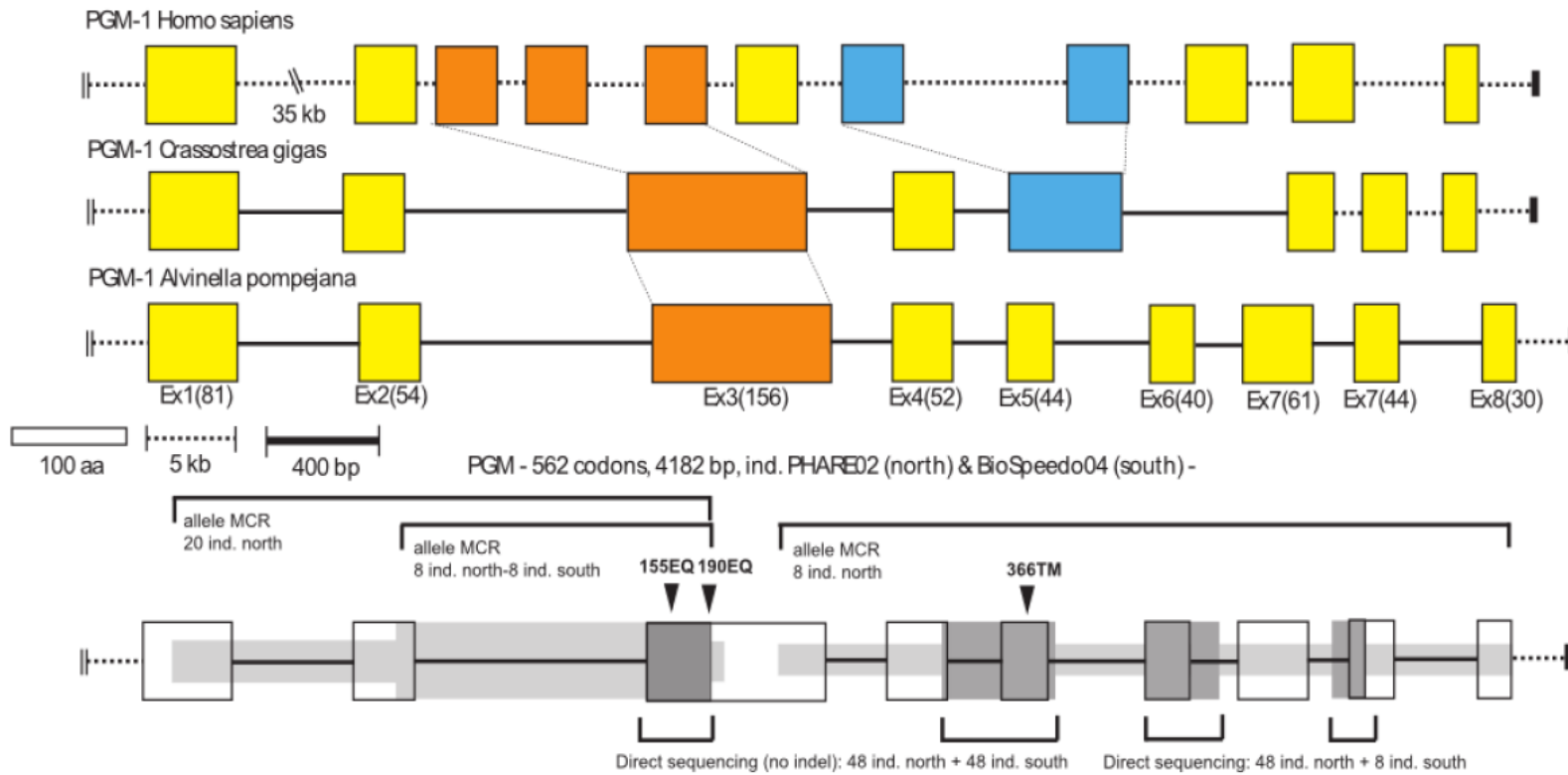


Fig. 3

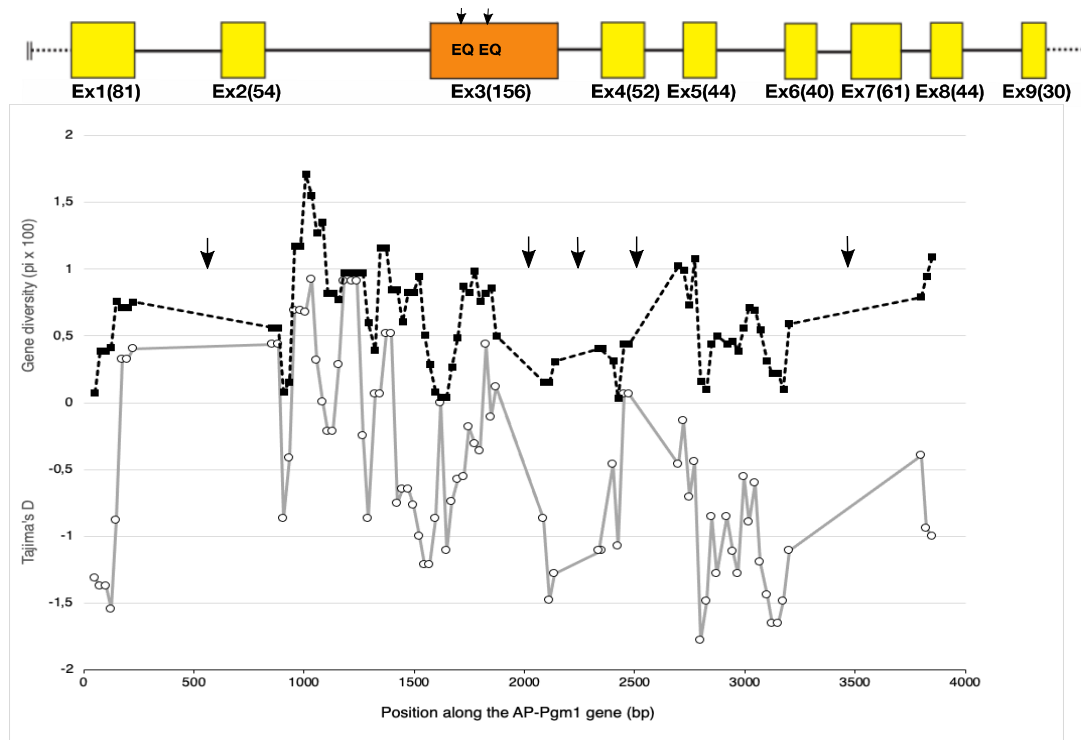


Fig. 4

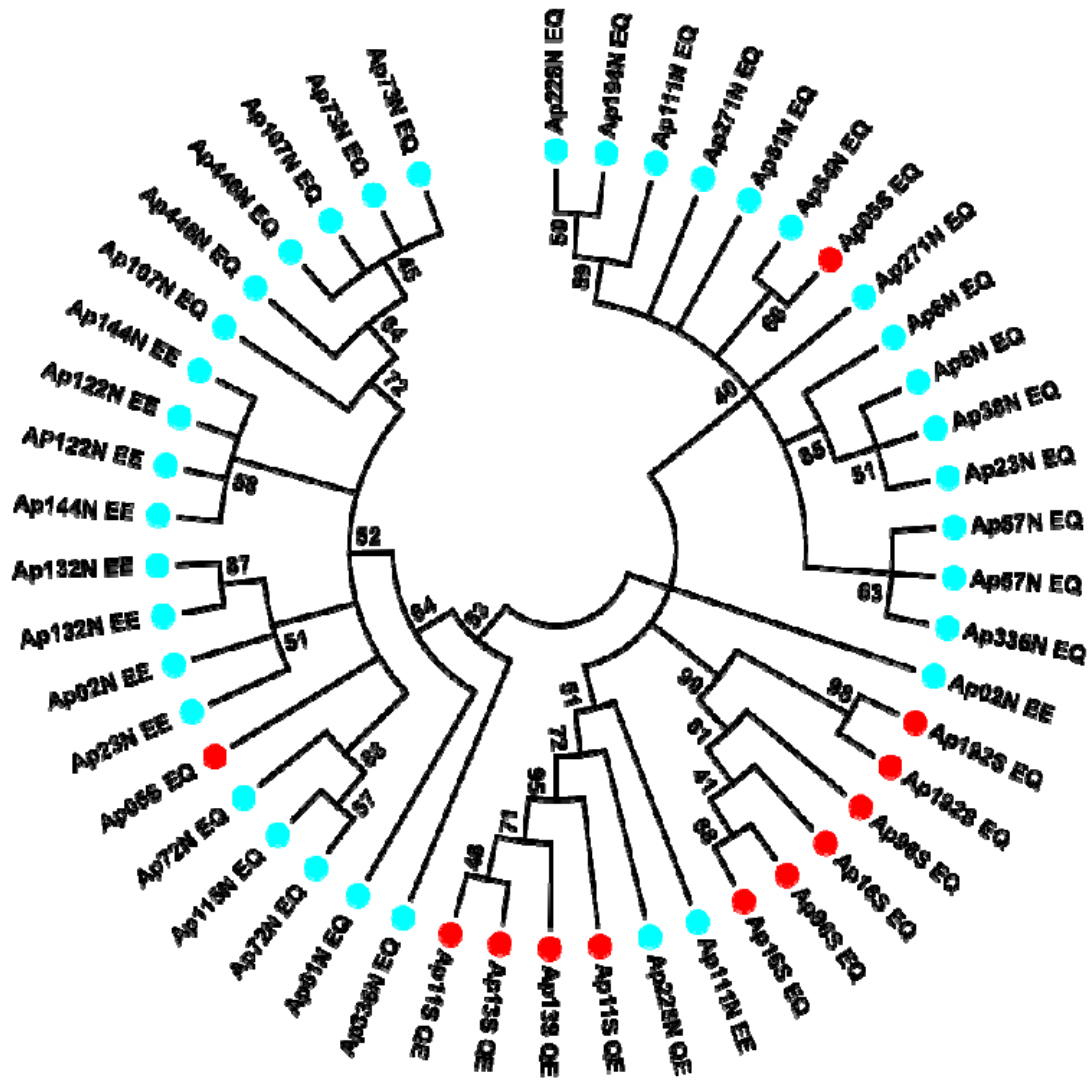


Fig. 5

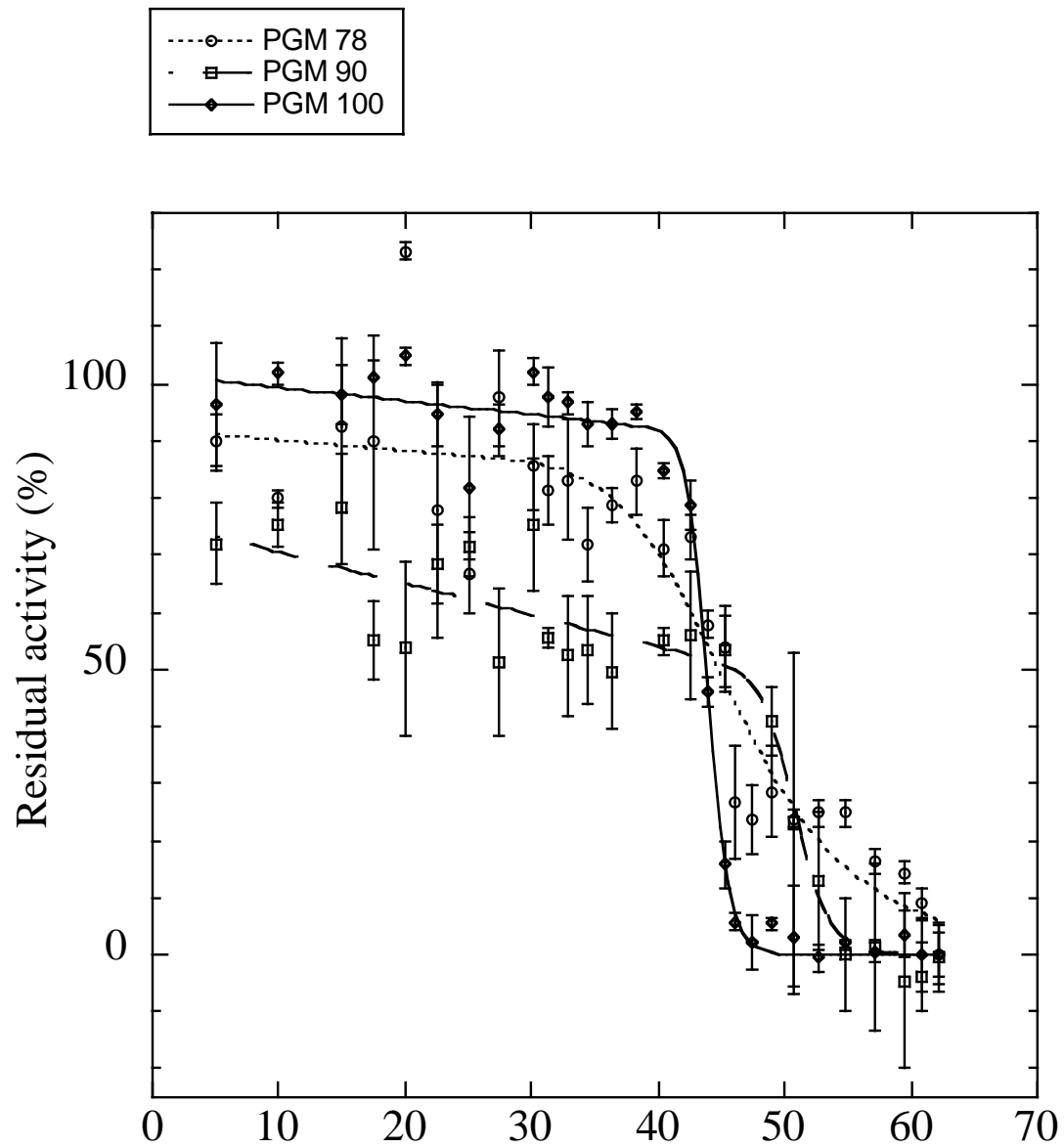


Fig. 6

