Article

The protected phosphoglucomutase polymorphism of the Pompeii worm *Alvinella pompejana* and its variant adaptability is only governed by two QE mutations at linked sites

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

[1] UMR 7144 Sorbonne Université-CNRS, Adaptation et Diversité en Milieu Marin, Equipe ABICE, Station Biologique de Roscoff, 29688 Roscoff, France

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

[3] Royal Belgian Institute of Natural Sciences, Rue Vautier 29, 1000 Brussels, Belgium. Email: mverheye@naturalsciences.be

[4] Lycée Lakanal, 3 Avenue du Président Franklin Roosevelt 92330 Sceaux, France. Email: patricepiccino@yahoo.fr

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

[6]. Current address UMR 8222 Sorbonne Université-CNRS, LECOB, Observatoire Océanologique de Banyuls, 66650 Banyuls-sur-Mer, France

*: corresponding author: Didier Jollivet, jollivet@sb-roscoff.fr, Tel: +33.(0)2.98.29.23.67, Fax: +33.(0)2.98.29.23.24

Abstract

The polychaete *Alvinella pompejana* lives exclusively on the walls of deep-sea hydrothermal vents along the East Pacific Rise. This environment is considered as extreme and highly variable and the worm displays specific adaptations to withstand high temperature and hypoxia. Previous studies revealed the existence of a balanced polymorphism on the enzyme phosphoglucomutase associated with differences in the thermal habitat of the worm. Allozymes 90 and 100 exhibited different optimal enzyme activities and thermostabilities. The exploration of the mutational landscape for allozyme variation of the phosphoglucomutase1 revealed the maintenance of four highly divergent allelic lineages that encode the three most frequent electromorphs, these alleles occurring at different frequencies in populations over the worm's geographic range. Enzyme polymorphism is only governed by two linked amino-acid replacements located in exon 3 (E155Q and E190Q). Unlike other studies dealing with the non-synonymous variations of the *Pgm* genes, these substitutions are not linked to other cryptic amino-acid polymorphisms. Overdominance under specific environmental 'hot' conditions should represent the most likely way for the long-term persistence of these isoforms. Using directed mutagenesis, overexpression of the three recombinant variants allowed us to test the additive effect of these two mutations on the biochemical properties of this enzyme. Results are coherent with those previously obtained from native proteins and reveal a thermodynamic trade-off between the protein thermostability and catalysis, which is likely to explain the long-term selection of these functional phenotypes before their geographic separation across the Equator with the emergence of a barrier to dispersal, about 1.2 Mya.

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

INTRODUCTION

A central goal in evolutionary biology is to understand the origin and maintenance of adaptations sculpted by natural selection, and more specifically how the mean phenotype of a population (i.e. the fundamental niche) can evolve under heterogeneous and/or changing conditions (Holt and Gaines 1992). The fundamental niche indeed corresponds to "the range of environmental conditions in which the mean fitness of a population exceeds or equals unity, and outside of which its value falls below one". As a consequence, most of adaptive improvements to local conditions usually fall within this range. However, when distinct populations exposed to distinct ranges of environmental conditions stay in contact, so when more than a niche are spatially available or when a niche changes through time, species can adopt a great variety of strategies. One of these strategies is to preserve adaptive polymorphisms by taking advantage on the level of habitat heterogeneity and the fitness of genotypes for each niche. The maintenance of differentially adapted forms to cope with the environmental variation in spatially heterogeneous environments thus remains a fundamental issue for adaptation of organisms (Schmidt *et al.* 2008). Most theoretical studies come from the two-niche models developed by Levene (1953) and Gillespie (1976, 1977) based on additive models of selection on enzyme polymorphism and subsequent habitat selection models (e.g. Templeton and Rothman 1981). In these models, the maintenance of a stable polymorphism mainly depends on the proportions of the two niches and the selection coefficient on genotypes (Hedrick 1986). However, the window for stabilized polymorphisms can often be challenged by the evolution of the niches through time depending on the amplitude of the temporal variations. If the change is cyclic or predictable, one observed strategy is the superimposition of different forms that are capable of coping with the environmental variation (Tauber *et al.* 1986). Alternatively, the species can enter resistance forms (i.e. diapause, cysts) until the return of the optimum conditions (Koornneef *et al.* 2002). Most of the time, long-term or seasonal fluctuations are likely to increase the niche size of the species through the phenotypic plasticity of the species. In the specific case of a two-niche model, this usually favours a loss of polymorphism when environmental effects are positively cross-correlated between the two niches (Hedrick et al. 1976). By contrast, short-term environmental fluctuations can extend the window of conditions for the maintenance of a protected polymorphism when they are negatively correlated between the two niches.

Many studies have already investigated the maintenance of enzyme polymorphism by

selective processes for species exposed to environmental gradients such as temperature, salinity or desiccation (Schmidt *et al.* 2008). Overdominance was first viewed as the simpliest mechanism for enzyme polymorphism protection but was not really adequate to explain levels of variation found in allozymes (Lewontin & Hubby 1966). As opposed, most of the documented cases for the long-term persistence of alleles by balancing selection, and trans-species polymorphism come from studies dealing with negative frequency-dependent selection at immune and sex-determination genes (Charlesworth 2006, Hedrick 2007). *In fine*, very few cases of protected enzyme polymorphisms have been reported with selection varying in space and time despite the convincing theoretical arguments brought up previously. This raises questions about how a chaotic and highly fluctuating two-niche system (abrupt and recurrent changes of niche proportions) can promote balancing selection at some enzyme loci.

Because of their tremendous thermal variability due to the chaotic mixing of the cold sea water and hot fluids, hydrothermal vents represent an ideal model for testing the effect of frequent -and unpredictable- spatial and temporal changes of habitats on 'adaptive' enzyme polymorphisms. The thermal and chemical nature of vent habitats mainly depends on the age of chimneys (Matabos *et al.* 2008). Vent fields often correspond to a mosaic of edifices of different ages (Jollivet 1996), whose mean age and size distribution is dictated by the frequency of tectonic and volcanic events and the dynamics of the heat convection beneath oceanic ridges (Shank *et al.* 1998, Marcus *et al.* 2003, Fontaine *et al.* 2008). Long-term oscillations of the heat convection indeed displace vents along the axial graben of the rift, generating the emergence of new vent sites more or less closely to older ones that became extinct (Watremez & Kervevan 1990, Jollivet *et al.* 1999). This has a profound impact on the metapopulation dynamics of the vent species (Vrijenhoek 1997, 2010) but also provides distinct ecological niches for pioneer species able to colonise chimneys at their very early stages. Still 'hot' (>120°C) diffusers indeed grow rapidly and cool down to become mature edifices such as snowballs, white smokers (if the inner fluid pipe is still porous) or black smokers (if the inner conduit is impermeable), which evolve slowly afterwards, from years to tens of years (Piccino *et al.* 2004).

This is especially the case of the polychaete *Alvinella pompejana*, which lives on the chimney hottest part of the environment (Chevaldonné *et al.* 1992, Cary *et al.* 1998) and can withstand temperatures up to 50°C (Ravaux et al. 2013). This tube-dwelling worm lives on the walls of hydrothermal-vent chimneys, from a latitude of 23°N on the East Pacific Rise

(EPR) to 38°S on the Pacific Antarctic Ridge (PAR) (Jang *et al.*, 2016) and has developed peculiar physiological adaptations to colonize this hostile habitat (Desbruyères et al. 1998, Chevaldonné et al. 2000). Earlier genetic studies already showed that the Pompeii worm presents a quite unusually high level of genetic diversity (Jollivet *et al.* 1995a, Plouviez *et al.* 2010, Jang et al. 2016) with a non-negligible number of bi-allelic enzyme loci with alleles at nearly the same frequency, some of which displaying different thermostabilities (Jollivet *et al.* 1995b). Amongst them, the enzyme phosphoglucomutase (PGM-1) exhibits four alleles, among which alleles 90 and 100 have intermediate frequencies of approximately 35% and 60% in the populations of the Northern East Pacific Rise, the two other alleles (112 and 78) being rather rare (accounting for the remaining 5%). Significant genetic differentiation was found by comparing contrasted microhabitats, especially the newly formed 'still hot' chimneys ('hot' niche) versus older and colder edifices ('cold' niche), with allele 90 being at its highest frequency in the former habitat. Moreover, the frequency of allele 90 was positively correlated with mean temperature at the opening of *Alvinella* tubes (Piccino et al., 2004). In parallel, thermostability and thermal optimum *in vitro* experiments showed that allele 90 was more thermostable and more active at higher temperatures than allele 100, and supported the view of an additive model of diversifying selection in which allele 90 is favoured in the 'hot' niche but not in the 'cold' one. However, *A. pompejana* populations also have a long history of transient gene flow and recurrent isolations due to geographic/tectonic barriers (i.e. transform faults/the Easter microplate), with possible gene introgression across these barriers (Plouviez *et al.* 2010, Jang *et al.* 2016). More precisely, Plouviez *et al.* (2010) showed that both PGM1 allozymes 78 and 100 displayed an abrupt clinal distribution across the Equator whereas the frequency of allozyme 90 remained constant over the species range. Together with these findings, the most interesting fact was that polymorphism seems to be preserved all along the EPR despite population isolation and a long history of divergence across the Equatorial barrier. This poses several questions about (i) the role of balancing selection in maintaining polymorphism within each group of populations at this specific locus and, (ii) the molecular processes responsible for the allozyme-specific thermostabilities and thermal performance.

Here, we report a possible case of long-term balancing selection at an enzyme locus where alleles can be maintained by an advantage towards heterozygotes under specific environmental conditions that greatly vary both in space and time. To locate adaptive mutations at the basis of the allozyme charge polymorphism and to demonstrate that these

mutations have a direct effect on the thermostability of the enzyme, we first sequenced the complete gene encoding the PGM-1 enzyme on individuals with distinct homozygous genotypes and compared exonic sequences between them. We then sequenced exon3 on worms previously genotyped using allozymes on either side of the equatorial barrier and showed that only two non-synonymous mutations fully explained the whole allozyme polymorphism of the worm. Full cDNAs containing these mutations were then used to produce recombinant enzymatic isoforms following directed mutagenesis on allele 100. Codon changes were only sought on the two polymorphic sites to test their potential effect on the catalytic and the structural properties of the PGM variants.

MATERIAL and METHODS

Animal sampling

Specimens of the Pompeii worm *Alvinella pompejana* were collected with the ROV Victor 6000 and the manned submersible Nautile on vent chimney walls at several locations of the East Pacific Rise (EPR hereafter) during the cruises Phare 2002 (chief scientist: N. Le Bris), Biospeedo 2004 (chief scientist: D. Jollivet) and Mescal 2010 (co-chief scientists: N. Le Bris and F.H. Lallier) on board of the N/O L'Atalante. The Phare 2002 and Mescal 2010 cruises targeted vent sites at 9˚50'N and 13˚N in the geographic range of the species, and the Biospeedo cruise targeted sites south of the equatorial barrier (South EPR, SEPR hereafter). Details about the sampling design and the sampled sites are given in Matabos et al. (2008) and Plouviez et al. (2010) for the two first cruises. Genotyped specimens were sampled over chimneys of different ages ranging from newly formed 'hot' diffusors to large black 'smokers', for which thermal and chemical conditions were highly contrasted (see: Jollivet et al. 2006, Matabos et al. 2008). Animal size was measured by the width of the S4 setigerous segment on board and the sex determined based on the presence of paired genital pores in females.

Fecundity determination

Mature females collected during the 2002 and 2004 cruises were dissected before genotyping for allozymes (see Faure et al. 2007 for more information about the female's oocytes counting procedure). For each female, coelomic fluid was carefully removed by making a ventral incision to the body wall of the trunk. The body cavity was extensively

washed out to collect all coelomic oocytes, which were then resuspended in 50 ml of a solution of borate-buffered 3% formalin in seawater. Oocytes (ranging from oogonia to mature oocytes) were counted from 2 oocyte suspension aliquots of 0.25 ml for each female.

Molecular analyses

Allozyme genotyping of PGM

A total of 220 individuals were genotyped for different parts of the gene (see Table 1 for details of number of individuals for each geographic zone and each part of the gene). Gill tissues were homogenized in an equal volume of a 0.01 M Tris-HCl pH 6.8 homogenizing buffer containing 2 mM EDTA, 0.5 mM NADP and 0.05% β-mercaptoethanol and 0.1mM PMSF. Homogenates were centrifuged for 30 min at 25 000 g and the supernatant was recovered for electrophoresis. Individuals were genotyped at the *Pgm-1* locus by electrophoresis on 12% starch-gel at 4°C (100 V, 80 mA, 4 h) with the Tris-citrate pH 8.0 buffer system following the procedure described in Piccino et al. (2004), and isoforms were chemically revealed following the PGM staining method 2 of Pasteur et al. (1987). In parallel, genomic DNA of the same individuals was extracted using a CTAB extraction procedure following the modified protocol described in Plouviez et al. (2009).

Amplification of cDNA fragments of Pgm-1 *using homozygous individuals*

Based on allozyme genotypes, 8 homozygous individuals for alleles 78, 90 and 100 were selected for RNA extractions. Total RNAs were extracted with Tri-Reagent (Sigma) following the manufacturer's instructions. A small piece of gill tissue was transferred to 500 µl of Tri-Reagent and ground with a TissueLyser MM300 Qiagen. Total RNAs were then separated by addition of 200µl of pure chloroform. Both quantity and quality of the RNAs and DNA were assessed through UV absorbance (OD 260/280/230) with a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Delaware, USA). Five µg of total RNAs were reverse transcripted with a M-MLV reverse transcriptase (Promega), an anchor-oligo(dT) primer (Table S1) and random hexamers (Promega). The reverse anchor and forward nested degenerated PGM primers derived from the oyster *Crassostrea gigas* and human *Pgm-1* sequences were then used to perform the upstream amplification of cDNA fragments from the Pompeii worm extracts (see Table S1). PCR-products containing *Pgm-1* cDNA candidates were then cloned with the TOPO TA Cloning kit (Invitrogen). Clones containing the target fragment were then sequenced on a ABI 3130 using BigDye v.3.1 (PerkinElmer) terminator chemistry following the manufacturer’s protocol. Sequences of clones containing the

appropriate *Pgm-1* cDNA fragments were then aligned to design specific reverse primers along the coding sequence. Using gDNA of the same individuals, these primers were then used to amplify the 5' portion of the gene by directional genome walking using PCR (Mishra et al., 2002). Briefly, the specific reverse primers (Table S1) were used in conjunction with a primer containing a random section. A nested PCR was then performed with a second nested primer and the non-random portion of the other primer. The products were cloned and sequenced following the same protocol as previously mentioned to determine which contained the sequence of interest. This allowed us to reconstruct a series of complete Pgm-1 cDNA including a reference sequence and three native consensus sequences of the three isoforms (Accession N° MN218832 - MN218839).

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

A series of specific primers were designed based on our cDNA sequence (see Table S1) to amplify both exon and intron-containing portions of the gene from gDNA on the same homozygous individuals. PCR amplifications were performed using pairs of the least distant barcoded forward and reverse primers along the gene in a 25 µl PCR reaction volume that included 1X buffer (supplied by manufacturer), 2 mM $MgCl_2$, 0.25 mM of each dNTP, 0.4 µM of each primer, 0.5 U of Taq polymerase (Thermoprime plus). The PCR profile included a first denaturation step at 94°C for 4 min followed by 30 cycles at 94°C for 30s, 60°C for 30s and 72°C for 2 min and, a final extension at 72°C for 10 min. Barcoded PCR-products were cloned following the MCR method developed by Bierne et al. (2007) and sequenced with the same protocol used previously. This allowed us to get the complete sequence of the *AP-Pgm-1* (Accession N° MN218831) and preliminary information on polymorphic sites between the 3 distinct genotypes all along the gene.

Genotyping Pgm-1 *exon 3 with RFLP and direct sequencing*

To examine the correspondence between the only two diagnostic polymorphic non-synonymous mutations EQ found at exon 3 and allozymes 78, 90 and 100, a 350 bp fragment of the *Pgm-1* exon 3 containing these sites was subsequently amplified by PCR using gDNA of the same 302 individuals that were previously genotyped with allozymes using specific genotyping primers (see Table S1). PCR amplifications were conducted in a 25 µl volume that included 1X buffer (supplied by manufacturer), 2 mM $MgCl_2$, 0.25 mM of each dNTP, 0.4 µM of each primer, 0.5 U of Taq polymerase (Thermoprime plus). PCR conditions were as follow: a first denaturation step at 94°C for 4 min followed by 40 cycles at 94°C for 30s, 60°C

for 30s and 72°C for 20s and, a final extension at 72°C for 2 min. After purification on a Nucleospin Extract II PCR clean-up column (following the manufacturer's instructions), PCR products were digested in a 25 µl reaction volume by either 0.25 µl of the enzyme Fai I (YA/TR targeting the first substitution site) at 50°C for 30 min, or the enzyme Bsg I (GTGCAG$(N)_{14}$/ targeting the second substitution site) for 2 hours at 37°C. RFLP genotypes were double checked with the direct sequencing of the PCR-products on ABI 3130 automatic sequencer with the BigDye v.3.1 (PerkinElmer) terminator chemistry after an ExoSAP-IT purification of the PCR products. Forward and reverse sequences were proof-read in CodonCode Aligner to check for the occurrence of single (homozygotes) or double (heterozygotes) peaks at the two polymorphic sites. The allele alignment has been deposited in Genbank (accession N° MN218918-MN219291). In addition, nucleotidic polymorphisms of exons 1, 4 and 5 were assessed by direct sequencing (Accession N° MN218840-MN218917 for exon1, Accession N° MN219292-MN219356 for exons 4 and 5) and that of intron 2 using the MCR method of Bierne et al. (2007) with a more restricted number of individuals (allele alignment with Accession N°. MN219357 - MN219404).

Plasmid construction for enzyme overexpression

To obtain full-length cDNA coding for the PGM, two homozygous individuals 9090 and 100100 were selected from samples coming from the north EPR (9°50'N on the BIO9 site during the cruise Mescal in 2010). RT-PCR was conducted with the ClonTech SMARTer Race cDNA amplification kit to obtain full-length cDNAs. First, the kit aimed at generating 5' RACE-ready cDNAs with a SMARTscribe Reverse Transferase (100 U) at 42°C for 90 min. Secondly, the whole *Pgm-1* cDNA was specifically amplified by 5'RACE PCR using the provided Long Up Universal primer and the AP-PGMex11 reverse primer (see Table S1). The PCR conditions were as follow: 25 cycles of 94°C for 30s, 68°C for 30s and 72°C for 3 min. These cDNAs were used as a target to specifically amplify the complete *PGM* coding sequence with two sets of primers specifically designed for the bacterial expression vectors Pet20 and PetDuet that contain the appropriate restriction sites to insert the cDNA in the relevant vector and either provides the start or the stop codon to the construction (Table S1). First, amplified cDNAs were cloned into a pGEM-t plasmid (PROMEGA pGEM-t Vector SystemTM kit) and DH5α competent *E. coli* cells were transformed with this construct. Minipreps of the constructs were then made (Qiaprep Spin Miniprep kit), and double-digested with the appropriate pair of enzymes (BamHI/NotI or AseI/XhoI) in a 25 µl volume

containing the restriction buffer, the enzymes, and 1% BSA. The fragments were separated on a 1.5% agarose gel, and the insert was purified with a Nucleospin Gel extraction Clean up column (Macherey Nagel). The restriction products were ligated in either the expression vector Pet20b or Pet-Duet. The obtained plasmid was used to transform BL21DE3 *E. coli* cells amenable for IPTG induction and overexpression.

Directed mutagenesis

Using the full-length cDNA with the double mutation EE as a template, the mutants 155EQ and 190EQ were produced by directed mutagenesis following Reikofski & Tao (1992)'s PCR protocol. This protocol is a two-step process: first, two fragments are created, one with the forward full-length primer and a reverse primer straddling the target mutated site, with a mismatch changing the codon, and the second with the forward primer (PGM-BamHI) with the mutated site (again straddling this mutated site) with the reverse full-length primer (PGM-NotI). These amplifications were conducted in 50 µl reaction volume containing: 1X Pfu buffer containing $MgCl_2$, 0.25 mM of each dNTP, 0.5 µM of each primer (petDuet and mutated primer), 0.5 U of the proof-reading Pfu polymerase (Promega) with 30 cycles of 94°C for 30s, 60°C for 30s and 72°C for 3 min. Secondly, the two regions of the mutated cDNA were joined following a new PCR amplification without primers mixing the two previous PCR-products (1:1) with the same conditions and a final elongation step of 10 min. The mutated cDNA 155EQ and 190EQ together with the native EE one were then sequenced on ABI 3130 automatic sequencer with the BigDye v.3.1 (PerkinElmer) terminator chemistry to verify the sequences before their overexpression in BL21DE3 *E. coli*, with the same protocol as described above.

Population genetics analyses

Basic genetic descriptors of the nucleotidic diversity of the *Pgm-1* gene were performed with the software DNAsp 4.10.3 5 (Librado et Rozas, 2009), based on the genotypes of a minimum of 8 individuals. This allowed us to estimate allele diversity Hd, nucleotide diversity π and its synonymous and non-synonymous components π_S and π_N, Watterson's theta θ_W, linkage disequilibrium between sites Z_nS, recombination rate R_m, and departures to neutral expectations (Tajima'D, Fu & Li's F, Fay & Wu's H) along the gene for the northern and southern EPR individuals. Genetic differentiation between the southern and northern parts of EPR was also estimated by calculating Fst and Dxy with the same software. Linkage disequilibrium between genotypes, EE, EQ and QE and allozymes 78, 90 and 100

was tested using the programme Linkdis (Black and Krafsur 1985) of the software Genetix v.4.05 (Belkhir et al. 2004). The exon 3 genotyping also allowed us to reconstruct an allele network using Network 4.5.1.0 (Bandelt et al. 1995), in order to examine the permeability of the equatorial barrier and test whether the southern and northern allelic lineages resulted from an ancestral polymorphism.

Functional and structural analysis

Protein expression and purification

E. coli (BL21DE3) transformed with the recombinant pETduet plasmid containing one of the native or mutated PGM cDNA sequences (PGM100, PGM90 and PGM78) were grown into a LB medium supplied with 100 µg/mL ampicillin at 37°C. When the cultures reached an absorbance of 0.6 at 600 nm, protein expression was induced by adding 1mM IPTG to the medium and kept at 37°C under shaking for 4 hours. Cells were then harvested by centrifugation (4°C/15 000 g/5 min), and the pellets were re-suspended in a binding buffer (20 mM Tris-HCl, pH 6.5, 500 mM NaCl, 5 mM imidazole), disrupted by French Press at 1.6 kbar and cell debris were removed by centrifugation (15 000 g/4°C/60 min). DNAse I (Eurogentec) was added to the supernatants (1 µg/mL of lysate) and the lysate was incubated for 1 hour on ice. A first purification step was then performed using immobilized metal affinity chromatography. Lysates were applied to His-bind resin column (His-Bond kit, Novagen) to recover PGM variants, through binding to the polyhistidine tail $(His)_6$ present in the pETduet plasmid. The columns (2 ml) were previously loaded with charge buffer (50 mM $NiSO_4$, pH 6.5) and then equilibrated with the binding buffer. After loading lysates, columns were washed twice with one column volume of binding buffer (20 mM Tris-HCl, pH 6.5, 500 mM NaCl) containing 60 mM and 5 mM imidazole and the PGM variants were finally eluted in a 20 mM Tris-HCl, pH 6.5, 500 mM NaCl, 1 M imidazole buffer. The eluted fractions were concentrated in 30 kDa molecular cut-off Amicon-Ultra (Millipore). A second purification step was performed by size-exclusion chromatography (SEC) with Superdex 75 column (1 x 30) (GE Healthcare) at a flow rate of 0.5 mL/min using a 25 mM Na_2HPO_4/NaH_2PO_4, pH 6.5. The protein elution was monitored at 280 nm. The purity of the products was checked by SDS-PAGE stained with Coomassie brilliant blue. The purified PGM were then conserved in the elution buffer supplemented with dithiothreitol (DTT, 10 mM) at 4°C until use for enzyme assays. The protein concentrations were measured by absorption at 280 nm with the theoretical coefficient of 48,820 $M^{-1}.cm^{-1}$ as calculated using the ExPASy-ProtParam tool (http://web.expasy.org/protparam/).

Enzyme activity assay

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

Thermal inactivation

The purified PGM activities were measured at 37°C under 340 nm using an UVmc^2 spectrophotometer (Safas, Monaco) after a 30-minute incubation at temperatures ranging from 5 to 60°C. Activities were then normalized as the percentage of residual activity when compared to the same sample kept in ice. A theoretical curve with the following equation was fitted to each experimental dataset using a nonlinear curve fit algorithm (Kaleidagraph 4.5.0, Synergy Software):

$$y = \frac{(yN+mN\,.T)+(yD+mD\,.T).\exp\left(\frac{m(T-Tm)}{RT}\right)}{1+\exp\left(\frac{m(T-Tm)}{RT}\right)}$$ (Pace and Scholtz *et al.* 1997) (1)

where y is the residual activity, y_N, m_N, y_D, m_D, the parameters characterising the activity of the native enzyme (N) and its denatured form (D), respectively, m characterising 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 reduce by half.

Guanidinium chloride-induced *unfolding of PGM isoforms*

Unfolding of the PGM isoforms was induced by guanidinium chloride (GdmCl) in a 25 mM sodium phosphate buffer, pH 6.5, NaCl 200 mM buffer. Proteins (12 µM) were incubated with increasing concentrations of GdmCl from 0 to 5 M, 30 min at 20°C and their intrinsic fluorescence emission was determined at 324 nm under excitation at 290 nm on a Safas Xenius spectrofluorimeter (Safas, Monaco). The GdmCl concentration was determined by refractive index measurements (Pace 1986). Biphasic states of protein denaturation with an intermediate state (I) between native (N) and unfolding (U) states according to the following

equilibrium: N↔I↔U were treated as follow: It was assumed that each transition (N↔I and I↔U) followed a two-state model of denaturation. The denatured protein fraction for each transition, f(I) for transition (N↔I) and f(II) for transition (I↔U), was determined by resolving the two following equations:

$$f(\mathrm{I}) = (y_N - y)/(y_N - y_I)$$
$$f(\mathrm{II}) = (y_I - y)/(y_I - y_U)$$

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

$$f(\mathrm{I\ or\ II}) = \frac{\exp\left(-m\frac{(Cm-[GdmCl])}{RT}\right)}{1+\exp\left(-m\frac{(Cm-[GdmCl])}{RT}\right)} \quad \text{(Walter \& Ron 2011) (2)}$$

where T is the absolute temperature, R is the universal gas constant, C_m is the concentration of GdmCl at the midpoint of the transition, m the dependence of the Gibbs free energy of unfolding reaction (ΔG) on the denaturation concentration of GdmCl. Knowing C_m and m, standard Gibbs free energy of the unfolding reaction in absence of denaturant, ΔG°_{H20}, can be calculated according to the relation:

$$\Delta G^{0}_{H20} = m\ C_m \text{ (Pace and Schloltz, 1997) (3).}$$

3D PGM Structure Modelling

PGM 78, 90 and 100 3D protein conformations were modelled with the Modeller 9v13 software (Sali & Blundell 1993), using the structure of the crystallized rabbit phosphoglucomutase with its substrate α-D-glucose 1-phosphate as a template (pdb file 1C47). This protein comprises 561 amino acids with a resolution of 2.70 Å that shares 65% sequence identity with that of *Alvinella pompejana*. One hundred models were generated for each PGM isoform and their quality was assessed using the Modeller Objective Function parameter. Finally, a structure optimization was obtained using the repair function of the FoldX software (Guerois et al. 2002). Local structural entropy (LSE) profiles were determined from the three PGM isoforms primary sequences using the web server http://sdse.life.nctu.edu/tw (Chan et al. 2004), with tetrapeptides and the SCOP-35 databases.

RESULTS

Effect of genotype on female *Alvinella pompejana* fecundity

Fecundity was estimated on 104 freshly captured females (see Faure et al. 2007). There were no heterozygous individuals with genotype PGM100/78. All other genotypes, whether homozygous (PGM100/100, 90/90 and 78/78) or heterozygous (PGM100/90 and 78/90), exhibit an average fecundity of 200,000 oocytes per female with a great variability among them (Faure et al. 2007). Female fecundity distributions were not significantly different one to each other according to genotype (Fig. 1).

Sequencing *Pgm-1* cDNA from homozygous genotypes

The amplification of full-length *Pgm-1* cDNA sequences was obtained from three genotypes 100/100, three genotypes 90/90 and only two genotypes 78/78. This led to a complete cDNA sequence of 562 codons without indel (including the initiation codon to the stop codon) for alleles encoding the three distinct allozymes (see cDNA sequences: accession N°MN218832-MN218839). The consensus protein sequence fell into the phosphoglucomutase 1 family of proteins with a blastp e-value of 0.0 (65-72% of identity over 99% of 562 residues with the sequence from the oyster *Crasostrea gigas*, and a range of vertebrate species). Out of the 16 clone sequences, only two non-synonymous mutations on exon 3 allowed us to discriminate three main genotypes (100/100, 90/90 and 78/78). These polymorphic mutations corresponded to the replacement of a glutamic acid (E) by a glutamine (Q) at positions 155 and 190. Another replacement of a valine (V) by a leucine (L) at position 40 was also found in exon 1 at intermediate frequency, but this amino-acid polymorphism was not linked with one of the three electromorphs. A phenylalanine (F) replacement by a leucine (L) was also found at position 502 in the cDNA encoding the allozyme 90. An additional amplification in the allozyme genotype 100/112 revealed a single clone with a replacement of an arginine (R) by an isoleucine (I) at position 342 in exon 4, which may represent the rarer electrophoretic isoform 112.

Pgm-1 gene structure characterisation

Using specific forward and reverse primers from the putative exonic regions (in reference to the *Pgm-1* human and the oyster *Crassostrea gigas* genes (Whitehouse et al. 1992, Tanguy et al. 2006), the structure of the gene was determined using genomic DNA from the same individuals. The full gene sequence with the location of polymorphic codons and primers are shown in FigS1. The total length of the *ApPgm-1* nucleotidic sequence is 4372 bp. The gene is subdivided into 9 exons and 8 introns which length ranges from 155 to 848 bp. The coding sequence of 1686 bp (562 codons) has an overall GC content of 43.5% (compared to only 29.3% in the introns). Introns are usually in phase 0 with the exception of intron 2, which is in phase 1, and introns 4, 5 and 6, which are in phase 2. When compared to the two other *PGM* genes already described in the literature, the largest third exon, comprising 156 codons (other exons vary from 40 to 81 codons), corresponds to the fusion of the exons 3, 4, and 5 of the human *Pgm-1*. This fusion of exons 3,4, and 5 has also been observed in the oyster *C. gigas* (see Fig. 2). This exon also displays potential cytosine methylation sites at the three last codons (cf. Fig S1).

Gene allele assignment to allozymes

Direct sequencing was performed on a small portion of exon 3 (94 codons) that contains the double diagnostic mutations EQ on 220 individuals from the Southern and Northern EPR. The linkage disequilibrium between the EQ mutations at codon positions 155 and 190, and allozymes was highly significant (Table 2) with correlation coefficients (Rij) greater than 70% (p-values<0.0001). This provides a very reliable correlation in which combinations QE, EQ and EE correspond to the isoforms 78, 90 and 100, respectively. The most negatively-charged allozyme 112, which is rare and always found at the heterozygous state in the northern populations was also assigned to genotype EE, suggesting that an additional replacement is occurring elsewhere in the protein.

Other -less common- variants not correlated to allozymes

Besides the main intermediate-frequency alleles that we linked to allozymic alleles, other, less common, variants at other positions in the protein were discovered in the natural populations. In the full-length *Pgm-1* cDNAs, additional mutations were more often encountered at the beginning of the gene on exon1 and on exons 4 and 5, The direct

sequencing of a gene fragment located between exon 6 and exon 8 did not show any diagnostic mutations between the 3 allelic lineages EE, EQ and QE, but genetic diversity was almost low and constant over the whole gene (Table 3).

Genotyping of exon 1 for individuals from the Northern and Southern populations also indicated the occurrence of a trans-equatorial V40L substitution found at a frequency of 0.15 restricted to the EQ (PGM90) allele for the Southern individuals and to only one of the two allelic lineages for the northern individuals, irrespective of the mutations EE (PGM100) and EQ (PGM90). The rare replacement of EQ at position 37 (only found in three distinct individuals) was also restricted in the southern population to the QE (PGM78) allele.

Frequencies of QE mutations in natural populations

The double mutation QE was genotyped (RFLP + direct sequencing) on exon 3 from a total of 220 individuals (Table 1). Groups of individuals from either the North or the South EPR did not depart significantly from the Hardy-Weinberg equilibrium. However, observed and expected heterozygosities were twice as high in the north than in the south (Ho_{-North}: 0.44 vs Ho_{-South}: 0.19). This value markedly varied from one locality to the other, depending on the chimney habitat. Interestingly, allele QQ was not found in any of the populations. Heterozygote deficiency observed at 17°25'S could be attributable to a Walhund effect with a clear increase of alleles EE in the population (also observed at 7°25'S: see Plouviez et al. 2010) but a significant excess of heterozygotes was also observed on the hottest young chimneys sampled in the study. The frequencies of EE, EQ and QE alleles at the different vent localities are summarized in Table 1. A more thorough analysis of the North/South genetic differentiation was conducted on 374 allelic sequences (see alignment in supplementary data). The resulting haplotype network (Fig. 3) shows a quasi-absolute isolation of these 2 populations with, again, a high Fst value again (Table 3). The PGM90 allele (EQ) found in the Southern population derives directly from the Northern one by one fixed mutation and the PGM78 allele (QE) differs by 2 mutations from the 100 allele (EE) mostly found in the North EPR populations.

Allelic lineages and divergence at the *Pgm-1* loci

The sequencing of the clones containing a part of the *Pgm-1* gene from the exon 2 to exon 3 including the two polymorphic sites EQ on 26 southern and northern individuals

previously assigned as 100/100, 90/90, 78/78, 100/90 and 90/78 highlighted the presence of a high level of synonymous polymorphism with a strong linkage disequilibrium between these sites, and two indels in intron 2 (insertions referred to as A and B following their order in the intron). At least, 9 and 11 additional synonymous mutations were fixed between allelic lineages in the southern (78 *vs* 90) and northern (L1 *vs* L2) individuals, respectively. In the Southern population, the sequence of allele 78 is typified by no insertion (QE, no_A, no_B) compared to allele 90 (EQ and A, B). The relationship between alleles is presented in figure 4 and shows that the two divergent lineages are in linkage disequilibrium with the 2 mutations E/Q without recombination. In the Northern population, two divergent lineages L1 and L2 also display linked sites with either the A indel (L1) or the B indel (L2) but these two lineages are not completely associated with the double mutations EE and EQ. For each lineage, alleles EE and EQ are respectively more frequent (*ca*. 75%) in one lineage, suggesting that these two alleles have recombined once very recently (Fig 4; Table 3). The direct sequencing of exons 4 and 5 also indicated that introns 3 and 4 were only present in the northern populations, at least in one of the two lineages, and thus reinforcing the hypothesis of a recombining event in the northern region (see alignments in Supplementary data). This study therefore indicated a very high divergence between these 4 allelic lineages, with divergence even greater between allelic sequences of the same population than those typifying the two geographic regions. With the exception of 2 sequences with a northern signature found in the Southern population, and a third one with a Southern signature retrieved in the Northern population (Fig. 3), allelic divergence comes with a significant high Fst value between these two metapopulations (Fig. 3, Table 3).

Comparative conformational stability and thermal inactivation of isoforms

To determine the conformational stability of the three recombinant isoforms of the PGM-1, their guanidinium chloride (GdmCl)-induced unfolding was studied. The variations of fluorescence emissions with increasing concentration of GdmCl were biphasic (Fig S2) suggesting two transitions and it was assumed that each of them following a two-state model of denaturation. For each transition, the unfolded fraction of protein (f_u) was determined (Fig S3) and Gibbs free energy change associated with each transition calculated (Table 4). For the two transitions, the PGM90 appears more stable than the two other isoforms. PGM78 appears more stable than PGM100 for the first transition ($\Delta G°_{H2O}$ = 8.0 ± 0.46 kJ.mol^{-1} *vs.* $\Delta G°_{H2O}$ = 6.06 ± 0.81 kJ.mol^{-1} respectively), whereas the conformational stability of the two isoforms is

not significantly different for the second transition ($\Delta G°_{H2O}$ = 15.43 ± 0.93 kJ.mol^{-1} *vs.* $\Delta G°_{H2O}$ = 15.13 ± 0.98 kJ.mol^{-1} respectively).

The T_m values obtained after adjustment of one theoretical curve on the thermic inactivation experimental data (inset Fig. 5) are very similar for PGM78 (46.5±1.7˚C) and PGM100 (44.0±0.1˚C), but that for PGM90 is markedly higher (50.9±0.7˚C).

Kinetic constants

The enzyme kinetic analyses of the three PGM isoforms were done at 37°C, a temperature close to the environmental conditions encountered by the animals (Table 4). The catalytic efficiency of the PGM78, evaluated by the ratio k_{cat}/K^{app}_m, appeared 125- and 70-fold higher than that of PGM90 and PGM100, respectively. Both of the two kinetic parameters, K_m^{app} (for substrate G1P) and k_{cat}, are involved in this important difference in catalytic efficiency between the three isoforms. Indeed K_m^{app} (G1P) and k_{cat} of the PGM78 are respectively tenfold lower and a tenfold higher than that of the two other isoforms. The PGM90 and PGM100 have a catalytic efficiency of the same order of magnitude.

DISCUSSION

A few decades ago, a series of enzymes interacting in the glycolytic cycle mostly associated with isomerase and mutase functions, such as GPI, MPI or PGM, have been shown to display isoforms that may be the subject of natural selection, leading to habitat-driven differentiation in populations according to temperature, wave action or metallic pollution (Nevo et al. 1977, 1983, 1986, Pogson 1989, Riddoch 1993, Schmidt & Rand 2001, 2008, Boutet et al. 2009, Bougerol et al. 2015). According to Eanes (1999), such branch-point enzymes, which are positioned at the crossroad of metabolic pathways, are likely to be the target of natural selection as they can orientate pathway flux according to their protein variation. Amongst them, alleles encoding the phosphoglucomutase enzyme have been widely studied with the aim of testing the hypothesis of differential and/or balancing selection by looking at allele (Nevo et al. 1977, 1983, Piccino et al. 2004) and heterozygote (Pogson 1991) frequencies in populations but also to assess either the fitness of individuals carrying suspected advantageous alleles among latitudinal clines (Verrelli et al. 2001a), or the kinetic properties of the allozymes themselves (Pogson 1989, Piccino et al. 2004). Although whole-length PGM sequences have now been obtained for a small set of metazoan species (Whitehouse et al. 1992, 1998, Verrelli & Eanes 2000, Tanguy et al. 2006), very few studies

have examined the kinetic effect of the non-synonymous changes at the gene level affecting the net charge of the PGM electromorphs (but see Verrelli & Eanes 2001a, 2001b for the correspondence between amino-acid changes and allozyme thermal resistance and glycogen storage in *Drosophila*).

Our group previously described a protected polymorphism at the *Pgm-1* gene in the northern populations of *Alvinella pompejana* (Piccino et al., 2004). The locus exhibited two equally-frequent isoforms (allozymes 90 and 100) and two rare ones (allozymes 78 and 112), and significant frequency changes were associated with variations of the thermal habitat of the worm (Jollivet et al., 1995a, Piccino et al., 2004). The two 'equally' frequent allozymes displayed distinct thermal stabilities and kinetic optima (Jollivet et al., 1995b, Piccino et al., 2004). In the present study, we sequenced the whole *Pgm-1* gene for the different alleles to assess the distribution of the non-synonymous polymorphisms along the gene and examine the relationship between these substitutions and the three main electromorphs described in natural populations. This allowed us to get a better understanding of the molecular basis of the enzyme adaptation to the highly fluctuating thermal regime of the vent habitat.

Structural effect of mutations E155Q and E190Q

The location of the two main polymorphic sites (E155Q and E190Q) onto the 3D model structure of the phosphoglucomutase 1 appeared exposed to the solvent and are not located in the binding domains of the enzyme (Fig. 6). Their potential effect on the catalytic properties of the enzyme is therefore not the result of a direct interaction with the substrate and/or the residues involved in the catalysis. This is not surprising as most of the mutations affecting the binding of glucose-1-phosphate, the ion Mg^{2+} and the phosphate should be deleterious. Similarly, in a study of polymorphism of the *PGM* from *D. melanogaster*, none of the 21 polymorphic amino-acid replacements were located in the catalytic site of the enzyme (Verrelli and Eanes 2001b). Based on their location, both the E155Q and the E190Q substitutions should affect the net charge of the protein in the same way. However, the isoforms 78 and 90 do not have the same electrophoretic mobility, suggesting that some post-translational modification may be involved in the electrophoretic separation of the three isoforms. The gain of a glutamate at position 155 may be associated with a potential ionic bond with histidine 157 with a distance of 6.5 Å between them. Ionic and hydrogen bonds have been shown to increase the stability of enzymes (Vogt et al. 1997), and partially explain the thermostable 3D structure of the Cu-Zn and Mn superoxide dismutase enzymes in *A.*

pompejana (Shin et al. 2009, Bruneaux et al. 2013). This may account for the increased thermostability of isoform 90. However, isoform 100 has the same substitution but its thermostability is not significantly different from that of isoform 78. This could suggest a negative effect of glutamate (E) at position 190 that would negate the positive effect of glutamate at position 155. The complete absence of the $Q^{155}Q^{190}$ combination also suggests a deleterious effect of these two replacements together (but not separately) in the native enzyme. The first mutation in position 155 is also likely to play a key role in the molecular dynamics of the protein, especially during the 180° rotation of the reaction intermediate, the glucose-1,6-diphosphate inside the active site. This likely explains the higher enzymatic efficiency of the isoform 78. On the other hand, the 3D model comparison of the three isoforms shows that the replacement of one glutamine (Q) by a glutamate (E) at position 190 (allozymes 78 and 100) introduces a negative charge in a region already enriched in acidic residues. This high density of negative charges could have a destabilizing effect on the protein structure by a Coulomb repulsion effect and thus could possibly lead to greater sensitivity to temperature. This hypothesis will be further explored by obtaining the crystalline structure of the three recombinant isoforms in order to compare and analyze their dynamics.

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

The non-synonymous polymorphism associated with the 4 allelic lineages of the *A. pompejana Pgm-1* was low (π_N=0.0037 on average). Such result sharply contrasts with earlier studies on *Drosophila melanogaster*, for which numerous cryptic non-synonymous changes have been described between the slow, medium and fast electrophoretic alleles of *Pgm-1* (21 non-synonymous polymorphisms; Verrelli & Eanes 2000, 2001a) or on the highly polymorphic phosphoglucose isomerase (PGI) alleles of *Colias* butterflies suspected to evolve under balancing selection (Wheat et al. 2005). For *A. pompejana*, only eight non-synonymous mutations (E37Q, V40L, E155Q, E190Q, R343I, G358S, T366M and F502L, some of which being at a low frequency) have been detected from the direct comparison of allelic sequences, and we found no specific increase of the nucleotidic diversity around the two selected sites. Analysis of the *Pgm-1* polymorphism clearly showed an allelic segregation of the northern isoforms $Q^{155}E^{190}$ (PGM78) and $E^{155}Q^{190}$ (PGM90) with the southern isoforms $E^{155}E^{190}$ (PGM100) and $E^{155}Q^{190}$ (PGM90) across the Equator, each of the four variants having accumulated a great number of synonymous substitutions since their separation. This allelic

segregation led to an overall significant F_{st} value of 0.588 between the two sides of the Equatorial barrier. The two allelic lineages found in the Southern populations exhibit 1.2% divergence over a fragment of 1110 bp located on intron2 and exon3. The Southern lineages correspond to the variants $E^{155}Q^{190}$ (PGM90) and $Q^{155}E^{190}$ (PGM78) with strong linkage disequilibrium between these mutations and the 11 silent substitutions found in intron2. This suggests that these two allelic lineages evolved separately without recombination (R_m=0) for a long period of time. The lineages are older than the divergence estimated between the southern and northern populations (average D_{xy}=0.009). The two other allelic lineages (L1 and L2) found in the northern populations diverge by 7 silent substitutions (0.6% divergence) and two diagnostic indels on intron2. These mutations are also linked, suggesting again that these two lineages evolved separately but for a shorter time. However, these diagnostic mutations were not exclusively associated with the variants $E^{155}E^{190}$ and $E^{155}Q^{190}$. Although one allelic clade corresponded to allozyme PGM90, the other one was a mixture of $E^{155}E^{190}$ and $E^{155}Q^{190}$ suggesting that at least one recombination event occurred between the two alleles or, alternatively, that the replacement of the glutamine by a glutamate at position 190 leading to allozyme 100 was recent and posterior to the equatorial separation of the Northern and Southern populations. The polymorphic presence of introns 3 and 4 in the northern allelic lineages L1 and L2 only but not in the southern ones also strengthen the hypothesis of a recombining event arising after the spatial isolation of populations across the Equator. According to the signature of the linked silent polymorphic sites of introns 1 and 2, both lineages of the Northern populations seem to derive from an ancestor of the Southern lineage associated with the $Q^{155}E^{190}$ genotype. In the haplotype network obtained from sequences of exon3, the Southern and Northern allelic lineages seem to have split simultaneously with the appearance of the physical barrier to dispersal, about 1.2 Mya. This clearly indicates that the co-occurrence of four highly divergent *Pgm-1* alleles derives from an older polymorphism predating the vicariant event that separated the Northern and Southern vent fauna of the East Pacific Rise, with the possible emergence of the genotype $E^{155}E^{190}$ in the North, secondarily. The co-occurrence of the amino-acid polymorphism V40L on the four allelic lineages also strengthens this scenario. However, even if the nucleotidic diversity was a little bit greater at the intronic region preceding the two selected sites, the analysis of the genetic diversity along the gene does not fit perfectly with the expectations of long-term balancing selection, with no specific 'hot spot' of silent site variation and significantly negative Tajima'D and Fu&Li'F tests near to the double selected sites $E^{155}Q/E^{190}Q$.

Selective modalities for the maintenance of a protected polymorphism

As opposed to other studies of the phosphoglucomutase, and of other branch-point glycolytic enzymes that control the metabolic flux for transport, storage and breakdown of carbohydrates (see Eanes 2011), selection does not seem to promote elevated levels of amino-acid polymorphisms in the deep-sea hydrothermal vent polychaete *A. pompejana*. This contrasts with theoretical effects of balancing selection on nearby genome regions promoting the increase of genetic diversity around the selected site due to both the lack of recombination and the long-term accumulation of mutations (Charlesworth, 2006). It appears instead to promote a temperature-driven balance of alleles based on only two polymorphic sites, $E^{155}E^{190}$ and $Q^{155}E^{190}$ variants being much more efficient at low temperatures and $E^{155}Q^{190}$ remaining active longer at higher temperatures. This agrees with the predictions of simple allozyme polymorphisms such as those found in 6PGD, G6PD or ADH in *Drosophila* (Eanes 1999) with two main electromorphs based on only one amino-acid substitution and a few rare variants. Such an unexpected behaviour of *A. pompejana Pgm-1* could however be partially explained by the very low level of enzyme polymorphism found in all vent species, whose population size vary greatly in time (frequent bottlenecks) and the challenging environmental conditions likely promote strong directional selection (both of which should favour enzyme monomorphism). Although we did not detect overdominance, the hypervariable nature of the vent habitat should favour overdominance, provided that the heterozygotes display fitness close to the favoured homozygote in one of the two habitats (Hoekstra et al. 1985). However, this can easily be masked by the dynamics of extinction-recolonization of populations living there and the temporal evolution of the thermal habitat with the age of the chimneys (hotter younger chimneys evolving to colder older chimneys).

The long-term evolution of *Pgm-1* alleles without recombination, at least in the first part of the gene, and their frequency changes according to environmental conditions raises questions about the selective modalities acting on the co-occurrence of alleles when one of the two alleles is better adapted to high temperatures (Piccino et al. 2004). So far, RFLP genotyping of the double mutation $E^{155}Q/E^{190}Q$ did not show any evidence for overdominance in populations with virtually no significant heterozygote excesses in the sampled populations. This sharply contrasts with the results reported by Pogson (1991) on the oyster *Crassostrea gigas*, in which overdominance seems to be the most likely evolutionary mechanism at the

origin of the maintenance of a protected polymorphism at the *PGM-2* locus (but also see Gardner & Lobkov 2005 for its link with individual's growth rate). In our case, the maintenance of allozymes with distinct thermostabilities may be more related to a two-niche dynamics model of local differentiation with habitat and drift (Levene 1953, Gillespie 1985, Hedrick et al. 1976, Hedrick 1986). However, *Fis* values in populations with distinct thermal habitat clearly indicated that populations exposed to hotter conditions displayed more heterozygotes than their colder counterparts, a situation that could be linked with the extreme variations of temperatures encountered at newly formed (and therefore hotter) chimneys. Indeed, animals can experience temperature amplitudes greater than 20°C within an hour (Le Bris & Gaill 2007). The alternate exposure to the cold and well-oxygenated sea water, on one hand, and the warm hypoxic fluid, on the other hand, allow the worm to store oxygen sequentially. Decoupling respiratory and nutritional physiologies might favour heterozygotes. This should not be the case for individuals living on older chimneys, where temperature variations are much milder, with values rapidly decreasing below 15-20°C, and varying little around these values.

Because the worm is also exposed to a mosaic of fluctuating thermal habitats where temperature could spatially vary according to the age of the chimneys (Piccino et al. 2004), a two-niche model of balancing selection may however be more appropriate to explain the maintenance of the *Pgm-1* polymorphism of the worm. Given this spatial and temporal dynamics of vents, the frequency changes of *Pgm-1* alleles could be either the result of local adaptation to the thermal heterogeneity of the sites or simply due to an exacerbated genetic drift associated with the dynamics of colonization of the newly opened sites. Indeed, the dynamic nature of hydrothermal vents over longer time scales (years) led to a very patchy habitat, scattered along the EPR with a complex heterogeneity of age-driven vent conditions that can be seen as a multitude of distinct ecological niches for the same species. In this context, the proportion of newly formed 'still hot' chimneys and older ones greatly varies over time depending on the spreading rate of the rift and thus the frequency of tectonic and volcanic events along the East Pacific Rise. Piccino et al. (2004) already proposed that the maintenance of a two-allelic polymorphism at the *Pgm-1* locus could be linked to the fact that the first settlers on newly-formed and still 'hot' (>100°C) anhydrite chimneys may display more thermoresistant alleles but could have a lesser fitness as measured by reproductive investment and/or survival. Watt (1977, 1983) and collaborators (1983, 1988, 2005) indeed demonstrated that both PGI and PGM have a great contribution on the male mating fitness of

the *Colias* butterflies, probably as the result of longer and more vigorous flight within the day. We however did not detect fitness differences between *Pgm-1* genotypes (using female fecundity as a proxy). This suggests that the predisposition to colonize still 'hot' chimneys is not compensated by a reduced reproductive success to avoid the fixation of the advantageous allele. However, if allozymes PGM78 and PGM100 play a key role in the storage of glycogen, one can easily imagine that there is a correlation between the catalytic efficiency of the enzyme and the fitness of the animal, and this could interfere with the pairing behaviour of the worm during reproduction.

In the fruitfly, the *PGM* locus is a quantitative trait for glycogen storage and hence, the ability to survive better to starvation (Verrelli & Eanes 2001b). PGM activity is indeed positively correlated to glycogen contents. In *Drosophila*, non-synonymous polymorphisms at *Pgm-1* play a non-negligible role in the regulation of the metabolic energy pool along latitudinal clines where the decrease of temperature is compensated by an increase of the enzyme activity. Populations of *Drosophila melanogaster* living at the highest (and thus coldest) latitudes possess PGM allozymes with a higher catalytic efficiency and a greater glycogen contents. According to the theory of metabolic flux, this can be an adaptive way of temperature compensation to maintain the same glycogen contents over the latitudinal gradient. In the case of *A. pompejana*, thermal differences could be great between colonists and reproducers. As a consequence, colonists subjected to longer periods of high temperature (and associated hypoxia) may be maladapted to produce and use glycogen reserves and, thus, are not able to rapidly invest into reproduction. On the contrary, secondary settlers arriving in a much cooler state of the chimney and thus, more frequently exposed to colder oxygenated waters, are more likely to use their glycogen reserves to massively invest in the production of gametes as previously shown by Faure et al. (2007). Colder conditions indeed seem to be a prerequisite for releasing fertilized eggs after pairing as embryos are not able to develop at temperatures greater than 15°C (Pradillon et al. 2001).

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

Thermal compensation may therefore be directly linked to different biochemical phenotypes that interact with the growth rate and reproductive effort of the worms. The theory predicts that enzyme activity variation associated with genotypes must be substantial to affect metabolic fluxes. As a result, enzyme polymorphisms in natural populations do not affect

metabolic fluxes (Hartl et al. 1985). Differences in both protein thermostability and catalytic efficiency however explain both local differentiation and latitudinal clines of *Pgm* allele frequencies in populations of the oyster *Crassostrea gigas* (Pogson 1989, 1991) or *Drosophila melanogaster* (Verrelli & Eanes 2001b). To test this hypothesis, one could measure the effect of non-synonymous mutations on the functional properties of the enzyme and its conformational stability and their effect on the population fitness. The three recombinant isoforms of the PGM-1 ($E^{155}E^{190}$ (PGM100), $E^{155}Q^{190}$ (PGM90) and $Q^{155}E^{190}$ (PGM78)) obtained by directional mutagenesis exhibit a two-step denaturation phase that suggests the occurrence of an intermediate stage that may be due to the dynamics of the molecule, and possibly the formation of homodimers. The replacement of the glutamate by a glutamine at position 190 increases the conformational stability and thermostability of the protein and confirms that PGM90 is the most thermostable isoform out of the three recombinant proteins. As discussed by Piccino et al. (2004), carrying this allele may be advantageous during the colonization of newly-formed chimneys whose surface temperature usually exceeds 50°C. Unexpectedly, the replacement of glutamate by a glutamine at position 190 also led to a decrease in the catalytic efficiency of the enzyme when compared to the variants $Q^{155}E^{190}$ (PGM78) and $E^{155}E^{190}$ (PGM100), these two latter isoforms having a k_{cat}/K_m ratio much greater (100 times for PGM78 and nearly two-fold greater for PGM100) when compared to $E^{155}Q^{190}$ (PGM90). The recombinant PGM90 also exhibits the lowest affinity for the glucose-1-phosphate substrate at 17°C. Such a trend was not reported in other invertebrate species subjected to balancing selection studied so far, for which no correlation has been reported between protein stability and enzyme activity (Pogson 1989, 1991, Carter & Watt, 1988, Verrelli & Eanes 2001b). Increased thermostability of a protein is often associated with a decrease in the flexibility of the molecule, and thus the dynamics of the enzyme reaction (Somero 1978, 1995). As a consequence, enzymes adapted to warmer environments should have a lower specific activity. Our results are in perfect agreement with these theoretical expectations and support the positive role of a thermodynamic trade-off between thermostability and catalysis as previously proposed by Eanes (1999) to explain the co-occurrence of alleles. This trade-off is probably due to the great thermal differences encountered by the worm within its habitat and the balance of local conditions at small spatial scales (over a few kilometres), which vary through time. Following such a scenario, PGM90 can remain stable for a longer period of time but is less efficient to either produce or consume the glycogen reserves of the worm than the two other isoforms. To this extent, the fact that the k_{cat}/K_m ratio of isoform PGM78 ($Q^{155}E^{190}$) is much higher than that of the isoform PGM100

($E^{155}E^{190}$) can explain why isoform 78 is slightly more frequent in the Southern populations (about 80%) than isoform PGM100 in the Northern populations (around 70%) and also fits with the hypothesis that this genotype $E^{155}E^{190}$ may have arisen more recently. The balance between allele frequencies from both sides of the Equator may be dictated by the selective coefficient attributed to each genotype as the direct reflection of the catalytic efficiency difference between PGM90 and its alternative isoforms.

References

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution, 16(1), 37-48.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Catch, F. (2004). GENETIX 4.0. 5.2, Software under Windows™ for the genetics of the populations. *University of Montpellier, Montpellier, France*.

Bierne, N., Tanguy, A., Faure, M., Faure, B., David, E., Boutet, I., … & David, P. 2007. Mark-recapture cloning: a straightforward and cost-effective cloning method for population genetics of single copy nuclear DNA sequences in diploids. Molecular Ecology Notes, 7(4), 562-566.

Black, W. C., & Krafsur, E. S. (1985). A FORTRAN program for the calculation and analysis of two-locus linkage disequilibrium coefficients. Theoretical and Applied Genetics, 70(5), 491-496.

Bougerol, M., Boutet, I., Le Guen, D., Jollivet, D., Tanguy A. (2015) Transcriptomic response of the hydrothermal mussel *Bathymodiolus azoricus* mediated by heavy metals is modulated by *Pgm* genotypes and symbiont content. Marine Genomics, http://dx.doi.org/10.1016/j.margen.2014.11.010.

Boutet, I., Tanguy, A., Le Guen, D., Piccino, P., Hourdez, S., Legendre, P., & Jollivet, D. (2009). Global depression in gene expression as a response to rapid thermal changes in vent mussels. Proceedings of the Royal Society of London B: Biological Sciences, rspb20090503.

Bruneaux, M., Mary, J., Verheye, M., Lecompte, O., Poch, O., Jollivet, D., & Tanguy, A. (2013). Detection and characterisation of mutations responsible for allele-specific protein thermostabilities at the Mn-superoxide dismutase gene in the deep-sea hydrothermal vent polychaete *Alvinella pompejana*. Journal of Molecular Evolution, 76(5), 295-310.

Cary, S. C., Shank, T., & Stein, J. (1998). Worms bask in extreme temperatures. Nature, 391(6667), 545.

Carter, P. A., & Watt, W. B. (1988). Adaptation at specific loci. V. Metabolically adjacent enzyme loci may have very distinct experiences of selective pressures. Genetics, 119(4), 913-924.

Chan, C. H., Liang, H. K., Hsiao, N. W., Ko, M. T., Lyu, P. C., & Hwang, J. K. (2004). Relationship between local structural entropy and protein thermostabilty. Proteins: Structure, Function, and Bioinformatics, 57(4), 684-691.

Charlesworth, D. (2006) Balancing selection and its effects on sequences in nearby genome regions. PLoS Genetics, 2, 379-384.

Chevaldonné, P., Desbruyères, D., & Childress, J. J. (1992). ... and some even hotter. Nature, 359(6396), 593.

Chevaldonné, P., Fisher, C. R., Childress, J. J., Desbruyères, D., Jollivet, D., Zal, F., & Toulmond, A. (2000). Thermotolerance and the 'Pompeii worms'. Marine Ecology Progress Series, 208, 293-295.

Desbruyères, D., Chevaldonné, P., Alayse, A. M., Jollivet, D., Lallier, F. H., Jouin-Toulmond, C., ... & Arndt, C. (1998). Biology and ecology of the "Pompeii worm"(*Alvinella pompejana* Desbruyères and Laubier), a normal dweller of an extreme deep-sea environment: a synthesis of current knowledge and recent developments. Deep Sea Research Part II: Topical Studies in Oceanography, 45(1-3), 383-422.

Eanes, W. F. (1999). Analysis of selection on enzyme polymorphisms. *Annual Review of* Ecology and Systematics, 30(1), 301-326.

Eanes, W. F. (2011). Molecular population genetics and selection in the glycolytic pathway. Journal of Experimental Biology, 214(2), 165-171.

Faure, B., Chevaldonné, P., Pradillon, F., Thiébaut, E., & Jollivet, D. (2007). Spatial and temporal dynamics of reproduction and settlement in the Pompeii worm *Alvinella pompejana* (Polychaeta: Alvinellidae). Marine Ecology Progress Series, 348, 197-211.

Fontaine, F. J., Cannat, M., & Escartin, J. (2008). Hydrothermal circulation at slow-spreading mid-ocean ridges: The role of along-axis variations in axial lithospheric thickness. Geology, 36(10), 759-762.

Gardner, J. P. A., & Lobkov, I. (2005). A test for overdominance at the phosphoglucomutase-2 locus in Pacific oysters (*Crassostrea gigas*) from B-New Zealand. Aquaculture, 244, 29-39.

Gillespie, J. H. (1977). A general model to account for enzyme variation in natural populations. III. Multiple alleles. Evolution, 31(1), 85-90.

Gillespie, J. H. (1976). A general model to account for enzyme variation in natural populations. II. Characterization of the fitness functions. The American Naturalist, 110(975), 809-821.

Gillespie, J. H. (1985). The interaction of genetic drift and mutation with selection in a fluctuating environment. Theoretical Population Biology, 27(2), 222-237.

Guerois, R., Nielsen, J. E., & Serrano, L. (2002). Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. Journal of Molecular Biology, 320(2), 369-387.

Hartl, D. L., Dykhuizen, D. E., & Dean, A. M. (1985). Limits of adaptation: the evolution of selective neutrality. Genetics, 111(3), 655-674.

Hedrick, P. W., Ginevan, M. E., & Ewing, E. P. (1976). Genetic polymorphism in heterogeneous environments. Annual review of Ecology and Systematics, 7(1), 1-32.

Hedrick, P. W. (1986). Genetic polymorphism in heterogeneous environments: a decade later. Annual review of ecology and systematics, 17(1), 535-566.

Hedrick, P. W. (2007). Balancing selection. Current Biology, 17(7), 230-231.

Hoekstra, R. F., Bijlsma, R., & Dolman, A. J. (1985). Polymorphism from environmental heterogeneity: models are only robust if the heterozygote is close in fitness to the favoured homozygote in each environment. Genetics Research, 45(3), 299-314.

Holt, R. D., & Gaines, M. S. (1992). Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. Evolutionary Ecology, 6(5), 433-447.

Jang, S. J., Park, E., Lee, W. K., Johnson, S. B., Vrijenhoek, R. C., & Won, Y. J. (2016). Population subdivision of hydrothermal vent polychaete *Alvinella pompejana* across equatorial and Easter Microplate boundaries. BMC evolutionary biology, 16(1), 235.

Jollivet, D. (1996). Specific and genetic diversity at deep-sea hydrothermal vents: an overview. Biodiversity & Conservation, 5(12), 1619-1653.

Jollivet, D., Desbruyères, D., Bonhomme, F., & Moraga, D. (1995a). Genetic differentiation of deep-sea hydrothermal vent alvinellid populations (Annelida: Polychaeta) along the East Pacific Rise. Heredity, 74(4), 376.

Jollivet, D., Desbruyères, D., Ladrat, C., & Laubier, L. (1995b). Evidence for differences in the allozyme thermostability of deep-sea hydrothermal vent polychaetes (Alvinellidae): a possible selection by habitat. Marine Ecology Progress Series, 125-136.

Jollivet, D., Chevaldonné, P., Planque, B. (1999). Hydrothermal-vent alvinellid polychaete dispersal in the Eastern Pacific. 2. A metapopulation model based on habitat shifts. Evolution, 53(4), 1128-1142.

Jollivet, D. and the Shipboard Scientific Party (2005). The BIOSPEEDO cruise: a new survey of hydrothermal vents along the South East Pacific Rise from 7°24S to 21°33S. InterRidge News, 13, 20-26.

Koornneef, M., Bentsink, L., & Hilhorst, H. (2002). Seed dormancy and germination. Current Opinion in Plant Biology, 5(1), 33-36.

Levene, H. (1953). Genetic equilibrium when more than one ecological niche is available. The American Naturalist, 87(836), 331-333.

Le Bris, N., & Gaill, F. (2007). How does the annelid *Alvinella pompejana* deal with an extreme hydrothermal environment? Reviews in Environmental Science and Biotechnology, 6(1-3), 197.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25(11), 1451-1452.

Marcus, J., Tunnicliffe, V., & Butterfield, D. A. (2009). Post-eruption succession of macrofaunal communities at diffuse flow hydrothermal vents on Axial Volcano, Juan de Fuca Ridge, Northeast Pacific. Deep Sea Research Part II: Topical Studies in Oceanography, 56(19-20), 1586-1598.

Matabos, M., Le Bris, N., Pendlebury, S., & Thiébaut, E. (2008). Role of physico-chemical environment on gastropod assemblages at hydrothermal vents on the East Pacific Rise (13 N/EPR). Journal of the Marine Biological Association of the United Kingdom, 88(5), 995-1008.

Mishra, R. N., Singla-Pareek, S. L., Nair, S., Sopory, S. K., & Reddy, M. K. (2002). Directional genome walking using PCR. Biotechniques, 33(4), 830-2.

Nevo, E., Noy, R., Lavie, B., Beiles, A., & Muchtar, S. (1986). Genetic diversity and resistance to marine pollution. Biological Journal of the Linnean Society, 29(2), 139-144.

Nevo, E., Shimony, T., & Libni, M. (1977). Thermal selection of allozyme polymorphisms in barnacles. Nature, 267(5613), 699.

Nevo, E., Lavie, E., & Ben-Shlomo, R. (1983). Selection of allelic isozyme polymorphisms in marine organisms: pattern, theory, and application. Isozymes, 10, 69-92.

Pace, C. N., & Scholtz, J. M. (1997). Measuring the conformational stability of a protein. Protein structure: A practical approach, 2, 299-321.

Pace, C. N. (1986). [14] Determination and analysis of urea and guanidine hydrochloride denaturation curves. In *Methods in enzymology* (Vol. 131, pp. 266-280). Academic Press.

Pasteur, N., Pasteur, G., Bonhomme, F., Catalan, J., & Britton-Davidian, J. (1987). *Manuel technique de génétique par électrophorèse de protéines*. Lavoisier, Paris.

Piccino, P., Viard, F., Sarradin, P., Le Bris, N., Le Guen, D., & Jollivet, D. (2004). Thermal selection of PGM allozymes in newly founded populations of the thermotolerant vent polychaete *Alvinella pompejana*. Proceedings of the Royal Society of London B: Biological Sciences, 271(1555), 2351-2359.

Plouviez, S., Shank, T. M., Faure, B., Daguin-Thiébaut, C., Viard, F., Lallier, F. H., Jollivet, D. (2009). Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South. Molecular ecology, 18(18), 3903-3917.

Plouviez, S., Le Guen, D., Lecompte, O., Lallier, F. H., & Jollivet, D. (2010). Determining gene flow and the influence of selection across the equatorial barrier of the East Pacific Rise in the tube-dwelling polychaete *Alvinella pompejana*. BMC evolutionary biology, 10(1), 220.

Pogson, G. H. (1989). Biochemical characterization of genotypes at the phosphoglucomutase-2 locus in the Pacific oyster, *Crassostrea gigas*. Biochemical genetics, 27(9-10), 571-589.

Pogson, G. H. (1991). Expression of overdominance for specific activity at the phosphoglucomutase-2 locus in the Pacific oyster, *Crassostrea gigas*. Genetics, 128(1), 133-141.

Pradillon, F., Shillito, B., Young, C. M., & Gaill, F. (2001). Deep-sea ecology: Developmental arrest in vent worm embryos. Nature, 413(6857), 698.

Ravaux, J., Hamel, G., Zbinden, M., Tasiemski, A. A., Boutet, I., Léger, N., ... & Shillito, B. (2013). Thermal limit for metazoan life in question: in vivo heat tolerance of the Pompeii worm. PLoS One, 8(5), e64074.

Reikofski, J., & Tao, B. Y. (1992). Polymerase chain reaction (PCR) techniques for site-directed mutagenesis. Biotechnology Advances, 10(4), 535-547.

Riddoch, B. (1993). The adaptive significance of electrophoretic mobility in phosphoglucose isomerase (PGI). Biological Journal of the Linnean Society, 50(1), 1-17.

Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology, 234(3), 779-815.

Schmidt, P. S., & Rand, D. M. (2001). Adaptive maintenance of genetic polymorphism in an intertidal barnacle: habitat-and life-stage-specific survivorship of MPI genotypes. Evolution, 55(7), 1336-1344.

Schmidt, P. S., Serrão, E. A., Pearson, G. A., Riginos, C., Rawson, P. D., Hilbish, T. J., ... & Grahame, J. W. (2008). Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. Ecology, 89(sp11).

Somero, G. N. (1978). Temperature adaptation of enzymes: biological optimization through structure-function compromises. Annual Review of Ecology and Systematics, 9(1), 1-29.

Somero, G. N. (1995). Proteins and temperature. Annual review of physiology, 57(1), 43-68.

Shin, D. S., DiDonato, M., Barondeau, D. P., Hura, G. L., Hitomi, C., Berglund, J. A., ... & Tainer, J. A. (2009). Superoxide dismutase from the eukaryotic thermophile *Alvinella pompejana*: structures, stability, mechanism, and insights into amyotrophic lateral sclerosis. Journal of Molecular Biology, 385(5), 1534-1555.

Tanguy, A., Boutet, I., Boudry, P., Degremont, L., Laroche, J., and Moraga, D. (2006). Molecular identification and expression of the phosphoglucomutase (PGM) gene from the Pacific oyster *Crassostrea gigas*. Gene, 382, 20-27.

Tauber, M . J., Tauber, C. A., & Masaki, S. (1986). *Seasonal adaptations of insects*. Oxford University Press, UK.

Templeton, A. R., & Rothman, E. D. (1981). Evolution in fine-grained environments. II. Habitat selection as a homeostatic mechanism. Theoretical Population Biology, 19(3), 326-340.

Verrelli, B. C., & Eanes, W. F. (2000). Extensive amino acid polymorphism at the *Pgm* locus is consistent with adaptive protein evolution in *Drosophila melanogaster.* Genetics, 156(4), 1737-1752.

Verrelli, B. C., & Eanes, W. F. (2001a). Clinal variation for amino acid polymorphisms at the *Pgm* locus in *Drosophila melanogaster*. Genetics, 157(4), 1649-1663.

Verrelli, B. C., & Eanes, W. F. (2001b). The functional impact of *Pgm* amino acid polymorphism on glycogen content in *Drosophila melanogaster*. Genetics, 159(1), 201-210.

Vogt G, Woell S, Argos P. 1997. Protein thermal stability, hydrogen bonds,and ion pairs. Journal of Molecular Biology, 269(4): 631–643.

Vrijenhoek, R. C. (1997). Gene flow and genetic diversity in naturally fragmented metapopulations of deep-sea hydrothermal vent animals. Journal of Heredity, 88(4), 285-293.

Vrijenhoek, R. C. (2010). Genetic diversity and connectivity of deep-sea hydrothermal vent metapopulations. Molecular Ecology, 19(20), 4391-4411.

Walter, P., & Ron, D. (2011). The unfolded protein response: from stress pathway to homeostatic regulation. Science, 334(6059), 1081-1086.

Watremez, P., & Kervevan, C. (1990). Origine des variations de l'activité hydrothermale: premiers éléments de réponse d'un modèle numérique simple. Comptes rendus de l'Académie

des sciences. Série 2, Mécanique, Physique, Chimie, Sciences de l'univers, Sciences de la Terre, 311(1), 153-158.

Watt, W. B. (1977). Adaptation at specific loci. I. Natural selection on phosphoglucose isomerase of *Colias* butterflies: biochemical and population aspects. Genetics, 87(1), 177-194.

Watt, W. B. (1983). Adaptation at specific loci. II. Demographic and biochemical elements in the maintenance of the *Colias* PGI polymorphism. Genetics, 103(4), 691-724.

Watt, W. B., Cassin, R. C., & Swan, M. S. (1983). Adaptation at specific loci. III. Field behavior and survivorship differences among *Colias* PGI genotypes are predictable from in vitro biochemistry. Genetics, 103(4), 725-739.

Wheat, C. W., Watt, W. B., Pollock, D. D., & Schulte, P. M. (2005). From DNA to fitness differences: sequences and structures of adaptive variants of *Colias* phosphoglucose isomerase (PGI). Molecular Biology and Evolution, 23(3), 499-512.

Whitehouse, D.B., Putt, W., Lovegrove, J.U., Morrison, K., Hollyoake, M., Fox, M.F., Hopkinson, D.A. and Edwards, Y.H. (1992) Phosphoglucomutase 1: complete human and rabbit mRNA sequences and direct mapping of this highly polymorphic marker on human chromosome 1. Proceedings of the National Academy of Sciences of USA, 89(1), 411-415.

Whitehouse, D.B., Tomkins, J., Lovegrove, J.U., Hopkinson, D.A., and McMillan, W.O., 1998. A phylogenetic approach to the identification of phosphoglucomutase genes. Molecular and Biology and Evolution, 15, 456-462.

Table S1. Sequences of the primers and their use. Primer names are based on the human *Pgm-1* exon number, but do not always correspond to the exon/intron nomenclature in *A. pompejana*, due to the comparative fusion of some exons (case of AP-exon3 with corresponds to exon2, exon3 and exon4 in human). Lower-case sequences correspond to intronic regions and upper-case sequences correspond coding regions.

Primers	**5' Sequence 3'**	**Localization/Method**
Anchor-OligodT	5-CTCCTCTCCTCTCCTC-T(17)-3	Primer for reverse-transcription of polyA+ mRNA
AP_PGMcDNAF1	5-GTNGTYGGNGGNGAYGGNBG-3	Pgm-1 cDNA fragment amplification
AP_PGMcDNAF2	5-YCAYAAYCCNGGNGGNCC-3	
AP_PGMcDNAR	5-NGTRATNACNGTNGGWKC-3	
Ap_PGMex1F	5-AAG AGG CAT CAG AGA AGA TA-3	Gene structure determination (amplification of exon1-intron1)
Ap_PGMex2R	5-GAC CAC CTG GGT TAT GAG AT-3	
Ap_PGMex2F	5-TAG GTA AAG ATG GCA TAC TT-3	Gene structure determination (amplification of exon2-exon3)
Ap_PGMex4R	5-CCG CTG AGA GCA TTG ACA AG-3	
Ap_PGMex5F	5-AA GAC TTT GGA GGA GGA CAT C-3	Chromosome walking along the gene (exon3-exon5)
Ap_PGMex7R	5-ATC CAT AAG GTT ACC AAA GAA CTT CCA-3	
Ap_PGMex7F	5-AGT TCT TTG GTA ACC TTA TGG ATG CT-3	Gene structure determination (amplification of exon5-intron6)
Ap_PGMex9R	5-ATT TTA TCA AGG TTG GCC ATC ATC TG-3	
Ap_PGMex9F	5-AC CTT GAT AAA ATG GCA GCT GAC-3	Gene structure determination (amplification of exon7-exon8)
Ap_PGMex10R	5-GA ATC TGA CTC ATA GCT ATC AAT GTA-3	
Ap_PGMex10F	5-ATG TAC ATT GAT AGC TAT GAG TCA GAT-3	Gene structure determination (amplification of exon8-exon9)
Ap_PGMex11R	5-AAA TTT AAG TAA TAA CAG TAG GCT GCT-3*	**anchored with the stop codon*
Ap_PGMint1R1	5-tgtgtatataagacgttctttttactgtgg-3	Chromosome walking to obtain the 5'UTR and the first exon with w1, w2, w3 and wc (Mishra et al., 2002)
Ap_PGMint1R2	5-tcatgtaaaattcttgatatacttacacc-3	
Ap_PGMex1Rn	5-GTA GCC TTT CTC AGA CCA CTA GT-3	
Ap_PGMint4F	5-tgttaggtagcatgcctcca-3	Genotyping EQ mutations in exon3
AP_PGMex5R	5-TCC TCC AAA GTC TTC CAG TG-3	
Ap_PGMex1F	5-AAG AGG CAT CAG AGA AGA TA-3	Genotyping intron1-exon2

Ap_PGMex2R	5-GAC CAC CTG GGT TAT GAG AT-3	
Ap_PGMex6F	5-ttt tag GAC AGA AAC ATG ATA CTT GGT-3	Genotyping intron4-exon5
Ap_PGMint7R	5-taatattac CT GAT GTG ATC TGA TCC-3	
Ap_PGMex9F	5-AC CTT GAT AAA ATG GCA GCT GAC-3	Genotyping exons 6 and 8
Ap_PGMex10R	5-GA ATC TGA CTC ATA GCT ATC AAT GTA-3	
AP_PGMmut78F	5-AAG TGA TAG ACT CTG TGC AGG ATT ATA TGG AC-3	Directed mutagenesis to produce allele 78 from allele 100
AP_PGMmut78R	5-TAG TCC ATA TAA TCC TGC ACA GAG TCT ATC AC-3	
AP_PGMmut90F	5-TTG ACA CAA AAG ATC ACA CAA TAC CAC ACT GTG C-3	Directed mutagenesis to produce allele 90 from allele 100
AP_PGMmut90R	5-AGG CAC AGT GTG GTA TTG TGT GAT CTT TTG TGT C-3	
Pet20_PGMXhoI	5-CTC GAG AGT AAT AAC AGT AGG CTG CTG TC-3	Cloning in Pet20 overexpression vector
Pet20_PGMAseI	5-ATT AAT GAG TCT GAA GTC GGT GAC AGT GGC T-3	
PetDuet_PGMBamHI	5-GGA TCC GAG TCT GAA GTC GGT GAC AGT GGC T-3	Cloning in PetDuet overexpression vector
PetDuet_PGMNotI	5-GCG GCC GCT TAA GTA ATA ACA GTA GGC TGC TGT C-3	

Table 1. Frequencies of EE, EQ and QE *Pgm-1* alleles, heterozygosities and Fis (*: significant with 1000 permutations) in northern and southern populations of *Alvinella pompejana*. Note: Allele QQ was not found in any of the populations.

Site	**N**	**EE**	**EQ**	**QE**	**Hobs**	**He$_{(n.b.)}$**	**Fis**
South EPR Overall	126	0.051	0.139	0.810	0.293	0.324	+0.094
Krasnov (21°33'S, hot)	23	0.044	0.217	0.739	0.434	0.413	-0.053
Bordreaux (21°25'S, hot)	30	0.033	0.100	0.867	0.267	0.242	-0.105
Fromveur (18°25'S, hot)	32	0.000	0.047	0.953	0.094	0.091	-0.033
Rehu Marka (17°25'S, cold)	41	0.110	0.195	0.695	0.390	0.472	+0.176
North EPR Overall	94	0.718	0.277	0.005	0.404	0.410	+0.014
Jumeaux (13°N, hot)	19	0.605	0.395	0.000	0.684	0.491	-0.410*
Julie (13°N, cold)	28	0.696	0.286	0.018	0.429	0.441	+0.028
Genesis (13°N, cold)	27	0.741	0.259	0.000	0.296	0.391	+0.246
Elsa (13°N, cold)	20	0.825	0.175	0.000	0.250	0.296	+0.159

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

Mutation	Dij	Rij	Khi2	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*

Table 3. Gene diversities, population parameters and neutrality tests along the *Pgm-1* gene for *A. pompejana* populations of the South and North EPR.

Statistics	E1 North	E1 South	E2-I2 North	E2-I2 South	E3 North	E3 South	E4-E5 North	E4-E5 South	E7-E9 North	E7-E9 South
N	24	24	66	37	78	110	20	45	91	8
H_d	0.89 ± 0.01	0.69 ± 0.01	0.96 ± 0.01	0.85 ± 0.04	0.62 ± 0.04	0.47 ± 0.04	0.84 ± 0.01	0.94 ± 0.00	0.95 ± 0.02	0.96 ± 0.02
Overall π	0.0039 ± 0.0007	0.0082 ± 0.0009	0.0054 ± 0.0005	0.0089 ± 0.0005	0.0030 ± 0.0003	0.0052 ± 0.0007	0.0038 ± 0.0076	0.0061 ± 0.0006	0.0062 ± 0.0005	0.0042 ± 0.0008
π_{silent}	---	---	0.0050	0.0073	0.0025	0.0039	---	---	0.0063	
π_s	0.0195	0.0024	---	---	0.0039	0.0168	0.0105	0.0213	0.0055	
π_n	0.0048	0.0043	---	---	0.0030	0.0027	0.0013	0.0017	0.0036	
θ_W (S)	0.0174 (20)	0.0129 (15)	0.0169 (50)	0.0083 (22)	0.0102 (15)	0.0102 (16)	0.0071 (10)	0.0127 (22)	0.0213 (55)	0.0047 (7)
Z_nS	0.0834 (1^B/1)	0.0143 ^{NS} (0^B/0)	0.0356 (4^B/1176)	0.1548 (29^B/231)	0.0038 (1^B/91)	0.1290 (15^B/105)	0.1847 ^{NS} (0^B/0)	0.0355 ^{NS} (0^B/0)	0.0206 (0^B/1275)	0.1430 (0^B/21)
R_m	1(RDP n.d.)	0 (RDP n.d.)	1 (RDP n.d.)	0 (RDP n.d.)	1 (RDP n.d.)	4 (RDP n.d.)	1 (RDP n.d.)	1 (RDP n.d.)	4 (RDP n.d.)	1 (RDP n.d.)
F_{st} Hudson, Slatkin & Maddison	0.183***		0.247***		0.588***		0.341**		0.016^{NS}	
D_{xy}	0.0074		0.0095		0.0101		0.0075		0.0055	
Tajima's D	-1.76^{NS}	-2.24**	-2.49**	0.22^{NS}	-1.87*	-1.27^{NS}	-1.655 ^{NS}	-1.553 ^{NS}	-2.31**	-0.43^{NS}
Fu & Li's F	-3.80**	-4.05**	-2.88*	0.43^{NS}	-2.58*	-0.65^{NS}	-1.732 ^{NS}	-2.98*	-4.52**	-0.50^{NS}

^{B}: still significant after a Bonferonni test.

(RDP n.d.): recombinant not detected using automated RDP and bootscan packages of RDP v.3.44.

Level of significance following permutation tests (1000 re-samplings): * <0.05, **<0.01, ***<0.001, ^{NS}: not significant.

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

	PGM 78		PGM 90		PGM 100	
	1rst transition	2nd transition	1rst transition	2nd transition	1rst transition	2nd 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
$\Delta G^0{}_{H2O}$ ($kJ.mol^{-1}$)	8.00 ± 0.46	15.43 ± 0.93	11.46 ± 2.03	18.10 ± 2.76	6.06 ± 0.27	15.13 ± 0.98
Tm (˚C)	46.5±1.7		50.9±0.7		44±0.1	
$K^{app}{}_{m}$ (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}/ $K^{app}{}_{m}$ ($sec^{-1}.M^{-1}$)	252.63 ± 17.06		2.0 ± 0.1		3.51 ± 0.49	

Figure captions

Fig. 1. Scatterplots of female *A. pompejana* fecundities according to their PGM-1 allozyme genotypes. Fecundities were determined on freshly collected females on board the research vessel and the Pgm-1 genotype was determined in the laboratory. The numbers above each category indicate the number of individuals.

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. Haplotypes network obtained by genotyping of the *Pgm-1* exon 3 network on 187 individuals coming from the North and the South EPR. The red and the yellow correspond to the populations of the North and the South, respectively. The numeric values are the numbers of sequences forming each haplotype. Mutations non-synonymous and the concerning amino acids are presents on the branches.

Fig. 4. Tree of K2P distance obtained by BioNJ method on 8 individuals from the North EPR and 8 from the South EPR locations sequenced by Mark-Cloning-Recapture (MCR) of the *Pgm-1* introns 2 and exon 3. The sequences corresponding to the PGM 78, 90 and 100 are respectively identified by the letters QE, EQ and EE traducing the polymorphism in position 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 and190 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+}.

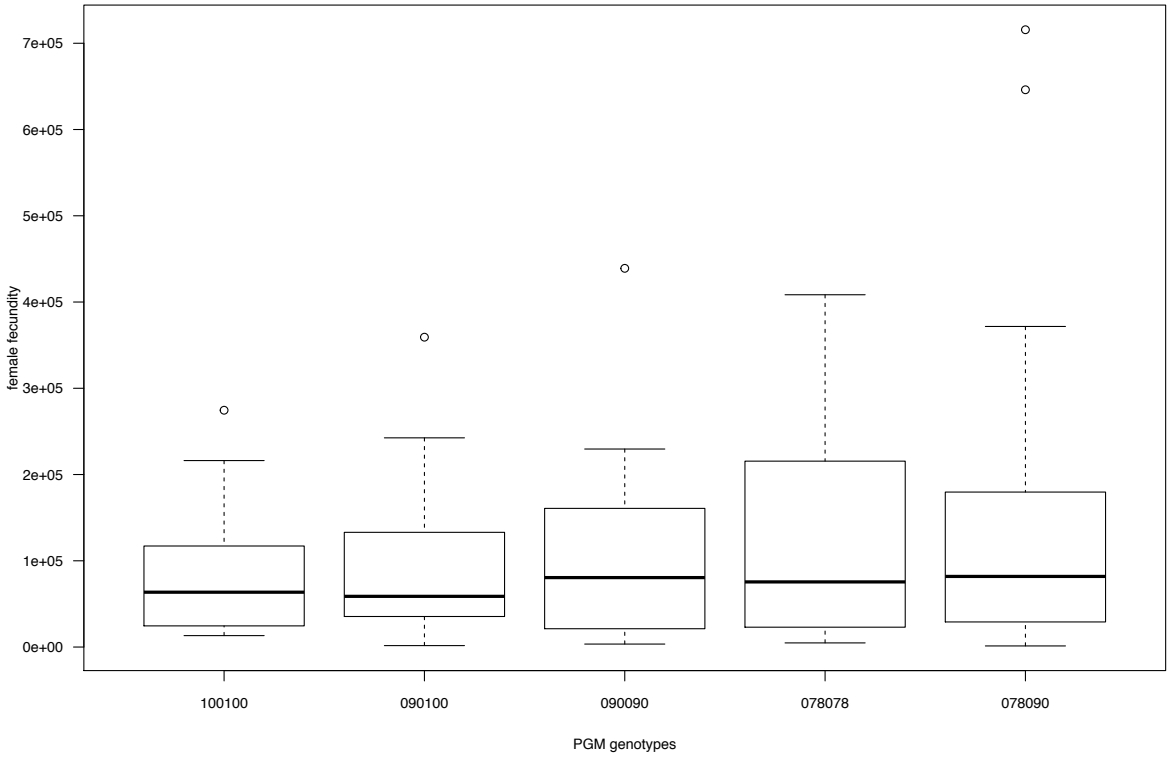

7e+05
6e+05
5e+05
4e+05
3e+05
2e+05
1e+05
0e+00
female fecundity
100100
090100
090090
078078
078090
PGM genotypes

Fig. 1

Fig. 2

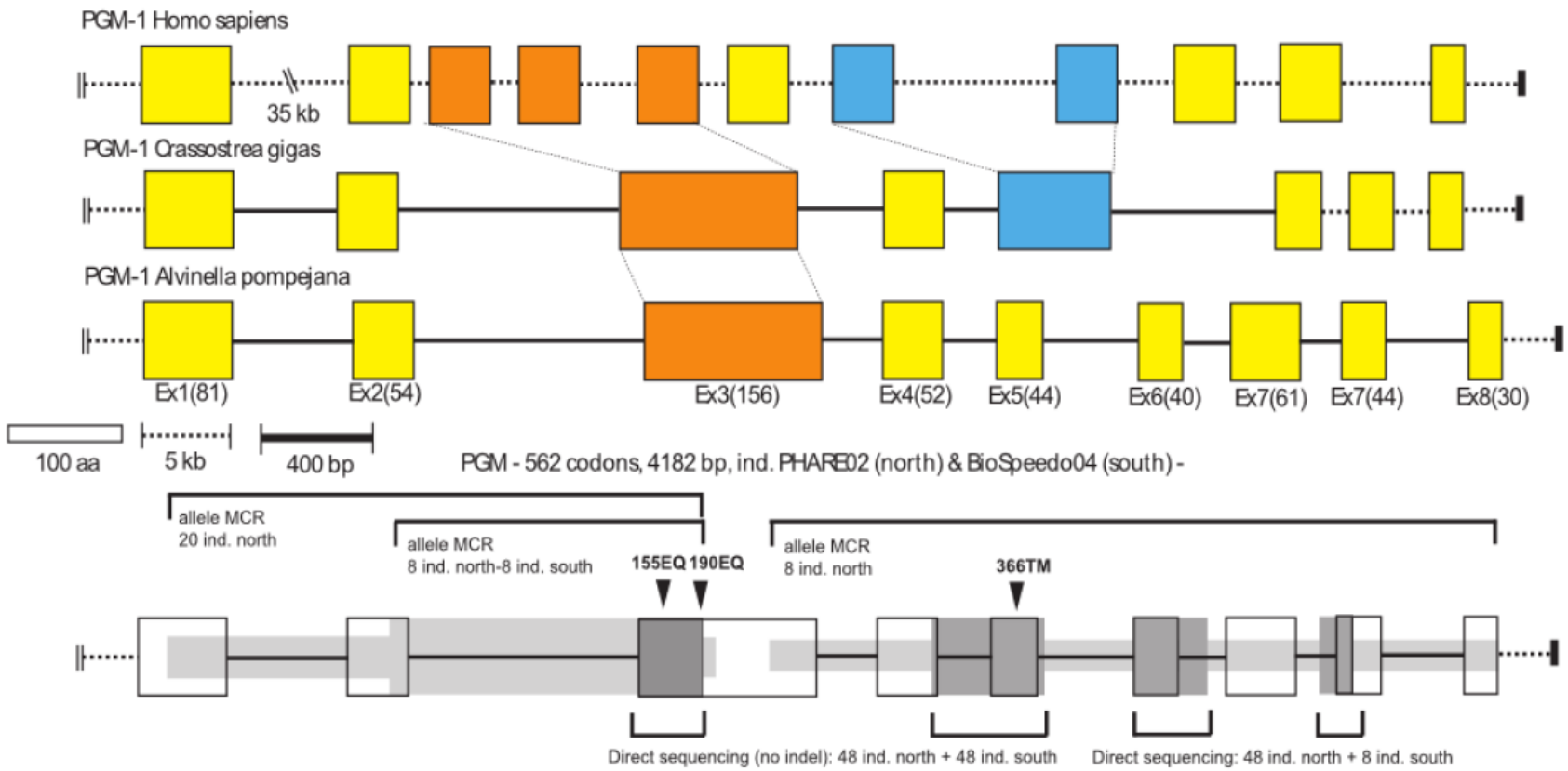
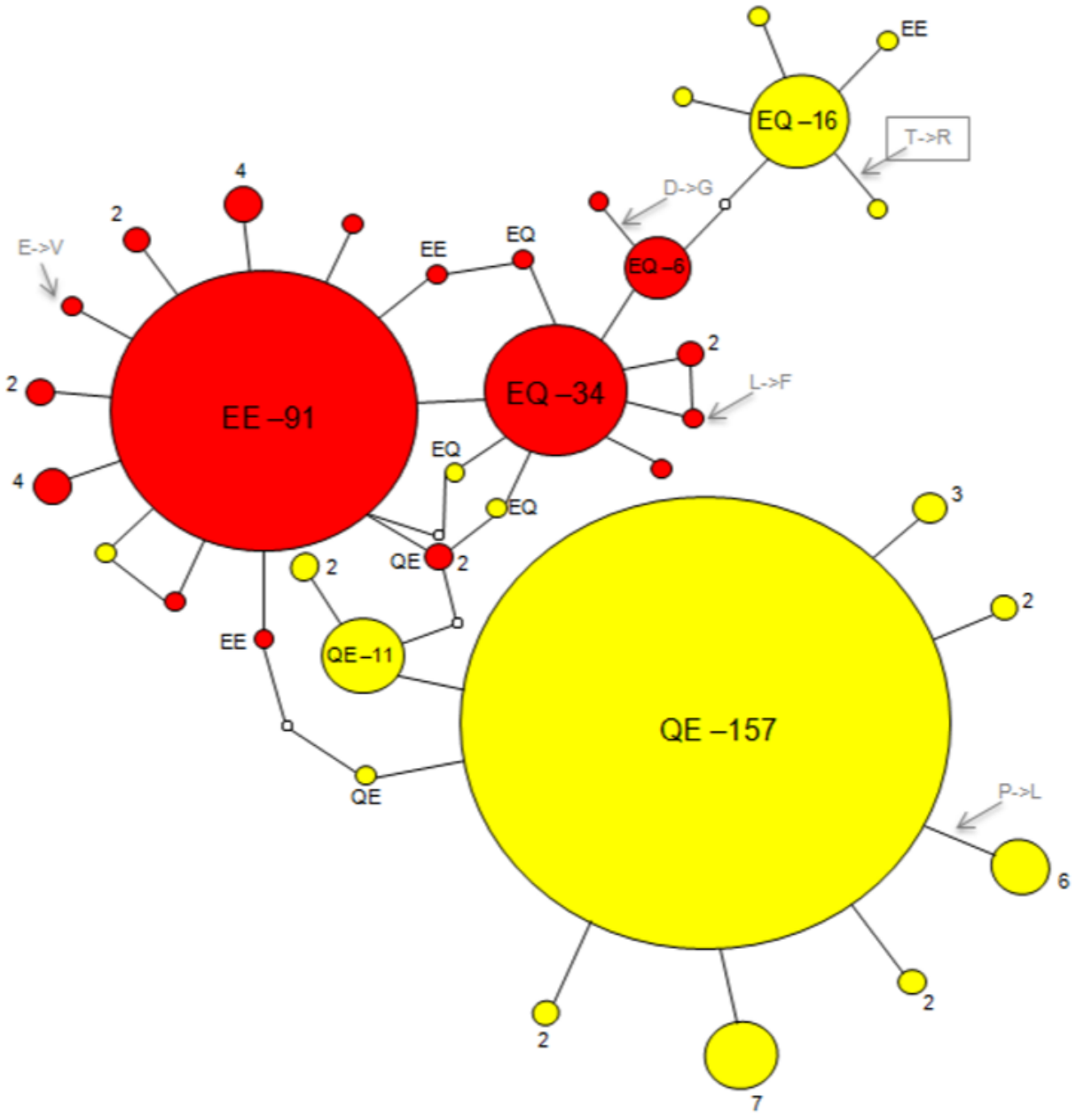

PGM-1 Homo sapiens
35 kb
PGM-1 Crassostrea gigas
PGM-1 Alvinella pompejana
Ex1(81)
Ex2(54)
Ex3(156)
Ex4(52)
Ex5(44)
Ex6(40)
Ex7(61)
Ex7(44)
Ex8(30)
100 aa
5 kb
400 bp
PGM - 562 codons, 4182 bp, ind. PHARE02 (north) & BioSpeedo04 (south) -
allele MCR
20 ind. north
allele MCR
8 ind. north-8 ind. south
155EQ 190EQ
allele MCR
8 ind. north
366TM
Direct sequencing (no indel): 48 ind. north + 48 ind. south
Direct sequencing: 48 ind. north + 8 ind. south

Fig. 3



EE
EQ –16
T->R
D->G
EQ –6
EQ
EE
4
2
E->V
EE –91
2
4
EQ –34
2
L->F
EQ
EQ
QE
2
2
QE –11
EE
QE
3
2
QE –157
P->L
6
2
2
7

Fig. 4

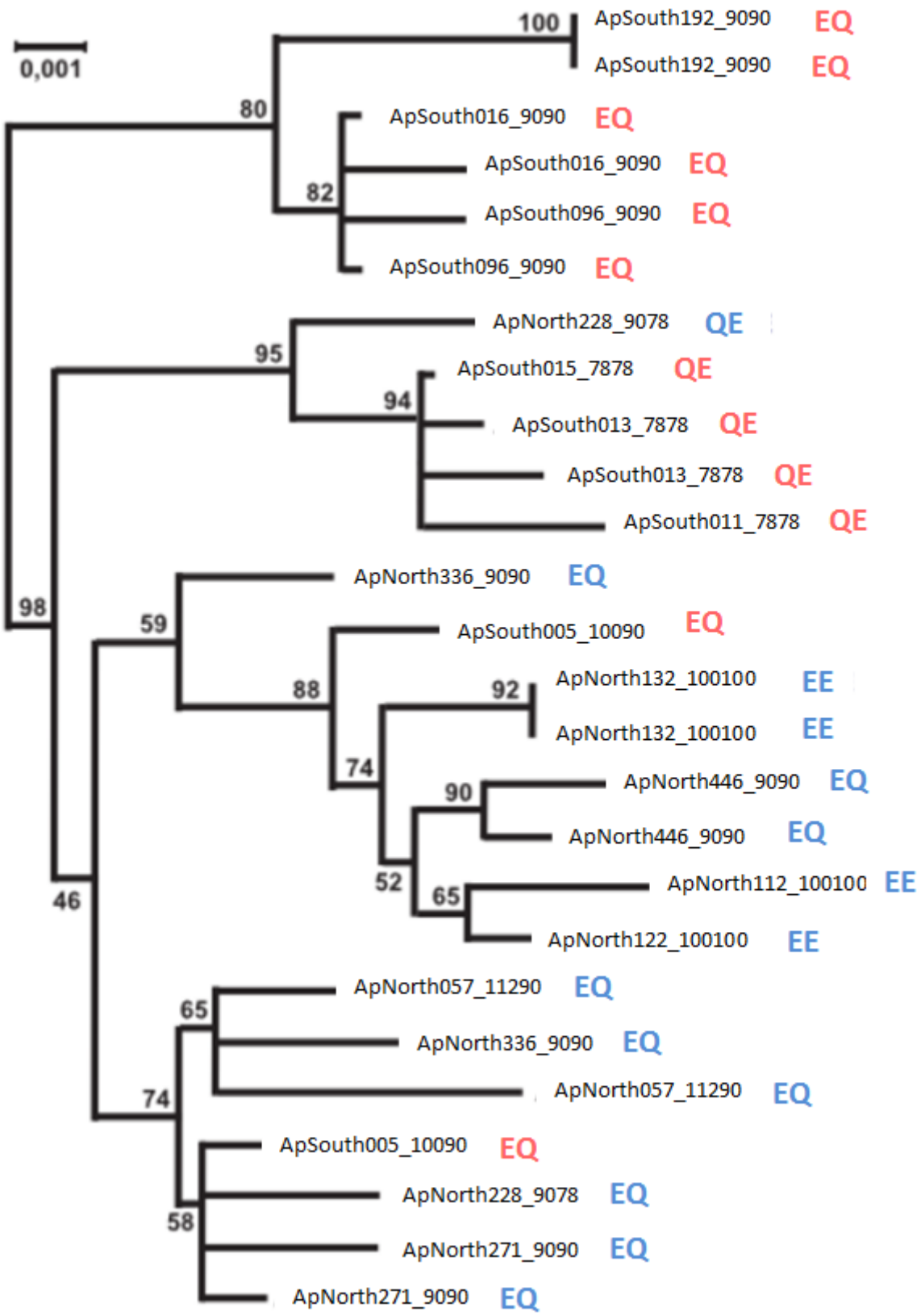

0,001
100
ApSouth192_9090 EQ
ApSouth192_9090 EQ
80
ApSouth016_9090 EQ
ApSouth016_9090 EQ
82
ApSouth096_9090 EQ
ApSouth096_9090 EQ
ApNorth228_9078 QE
95
ApSouth015_7878 QE
94
ApSouth013_7878 QE
ApSouth013_7878 QE
ApSouth011_7878 QE
ApNorth336_9090 EQ
98
59
ApSouth005_10090 EQ
88
92
ApNorth132_100100 EE
ApNorth132_100100 EE
74
90
ApNorth446_9090 EQ
ApNorth446_9090 EQ
52
65
ApNorth112_100100 EE
46
ApNorth122_100100 EE
ApNorth057_11290 EQ
65
ApNorth336_9090 EQ
74
ApNorth057_11290 EQ
ApSouth005_10090 EQ
ApNorth228_9078 EQ
58
ApNorth271_9090 EQ
ApNorth271_9090 EQ

Fig. 5

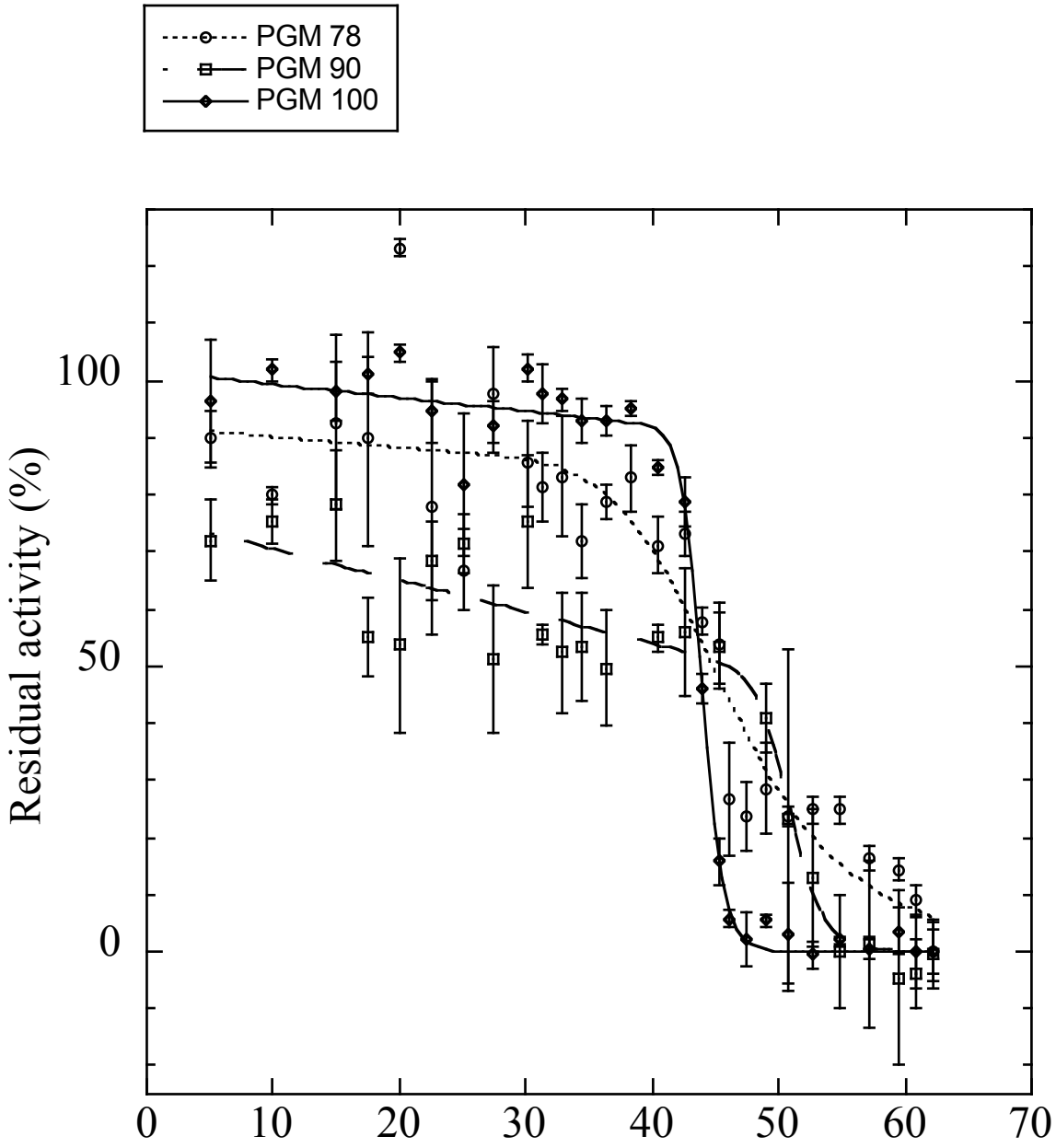

PGM 78
PGM 90
PGM 100
Residual activity (%)
100
50
0
0
10
20
30
40
50
60
70

Fig. 6

FigS1. Entire sequence of the *Pgm-1* gene of *A. pompejana* and its translated exons

(1) Structure of the *Pgm-1* gene. Grey zones represent positions where forward and reverse primers have been designed. Exons are indicated by the use of uppercase and introns are in lowercase. Highlighted codons in yellow* represent polymorphic non-synonymous changes between alleles found at a frequency of more than 10%, and in green below 5% but which are likely to change the net charge of the protein. # symbol represents methylated codons.

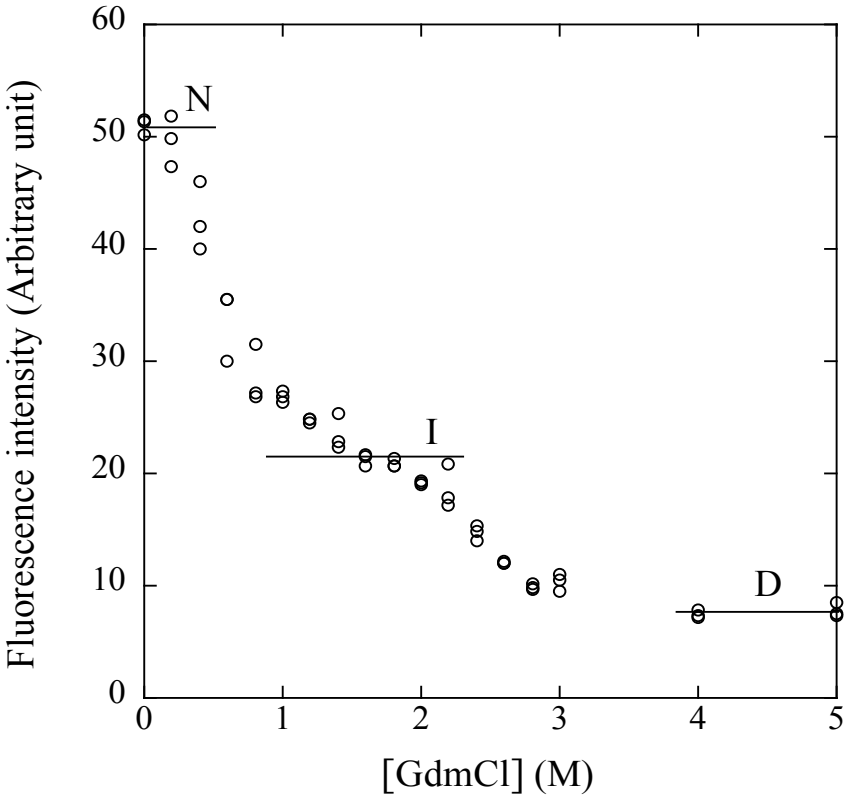
```
ATG AGT CTG AAG TCG GTG ACA GTG GCT ACG AAG CCC TTC GAT GGG CAG
AAG CCG GGC ACT AGT GGT CTG AGA AAG GCT ACG AAG ATA TTT ATG CAA
GAA CAT TAC ACA GAA AAC TTC GTG CAA TGT ACG TTG TCT GCC ATG GGC
GAC AAA TTA AAG GGA TGT ACA CTA GTA GTT GGA GGT GAT GGA AGG TAT
TAT GGT AAA GAG GCA TCA GAG AAG ATA ATT AAA ATG TGC GCA GGT AAT
GGT gtaagtatatcaagaattttacatgatgtgtaaac
aattgtttctaactgacatcagaagccacagtaaagagaacgtcttatatacacagataatata
tgttaacacaactttcttctcgttgcattaaatgcggtgattaatttttgctttatttgaaact
accaagtttatttaacgtatttatttgttcacaatttgaaatatgtttatactgtttacatgcc
tgaatttggtttgaattgacgtttttaaatgtactaattacgttttgttgttgttgtttttta
gactatttgaatcatttaaccagccatgctaattttgattctattgtttgtgcttattacattt
gtaccagatatgaaaggggagttaagaatgtgttgtaccatgatttgtatggtgatttaacatt
aatgcaattatgtttgtgtatttatatatag GTA GCA AAG GTA ATT ATA GGT AAA
GAT GGC ATA CTT TCT ACA CCA GCT GTG TCA TGC TTG ATC AGA AAA AAT
CAC ACT GAT GGA GGA ATA ATC CTC ACT GCA TCT CAT AAC CCA GGT GGT
CCA AAT GCT GAT TTT GGC ATA AAG TTT AAT ATT GCC AAC GGA G
gtaattcattcatgtttc
tacctgttcaaatcctttaaaccatacaaacaaatccttctatttgcaagaccaataaaaatgt
tcaatgtttaatcaatggtaacagctgaatgtaagtgtatgtatttttcaaatgtgatcaatat
aatataaatggactattttaaagacagtagagtctttctatgattgccatgtttttaattgatt
ctgtctttttaaccatctgttgtattaaggggtgaaactagtattctgcataatgctatatgtt
tattcattatctgttagatttgaagagaataatatcataatagttagttaacctttttattttt
agtagtgtttgtttcattaaaatggcacatcagctcattttgtagtgatttttatcatttgtgt
agtttatatttattaaagataaataatactctctacttttattagttgtagttgttttgggggt
ctatacaccaaactgttcatatcatgataactttatgactatgggtgttatgtcaccatgatag
aaacttgatatttaaagagcaaagtgaactagtagatccagattccacagttggtcctggttaa
aatattcaacactgtctgtcactgatgtacacaaattgttgtgatccatggatgttctataaat
gtttgcttagcatttgatgtaatatatgctacatagtttttgctgtatctactgcctgtccttt
tccatattttattgggactacttggtcactttagctgaccaaggacttgttaggtagcatgcct
ccattatttgtatgattaatgtaggtttgaaagctgaccaatgttaatatttttcatcacag
GA CCA GCC CCA GCT GGA GTC ACA GAT CAC ATC TAT GCC TTG ACA CAA
AAG ATC ACA GAA TAC CAC ACT GTG CCT GAC CTG AAG GCT GAC ATC TGT
ACA ATA GGA AGC CAG ACG TTT ACT GTT GAT GAT CAT CCA TTT AAT ATT
GAA GTG ATA GAC TCT GTG GAG GAT TAT ATG GAC TAC ATG AAG GAA ATT
TTT GAC TTC AAT TCC ATC AGA GGC TTA TTG ACT GGA GAA GGA GGA CAA
ACA AAG CTA AAG GTC CTT GTC AAT GCT CTC AGC GGA GTG GTT GGC CCA
TAT GTG AAG AGG ATA CTG TGC CAG GAG CTA GGC ATG GAT GAA GCC AGT
GCT GTT AAT TGT GTT CCA CTG GAA GAC TTT GGA GGA GGA CAT CCA GAT
CCC AAC TTG ACC TAT GCA GCT GAT TTA GTG AAT GAA TTA AAG AAA GGT
```

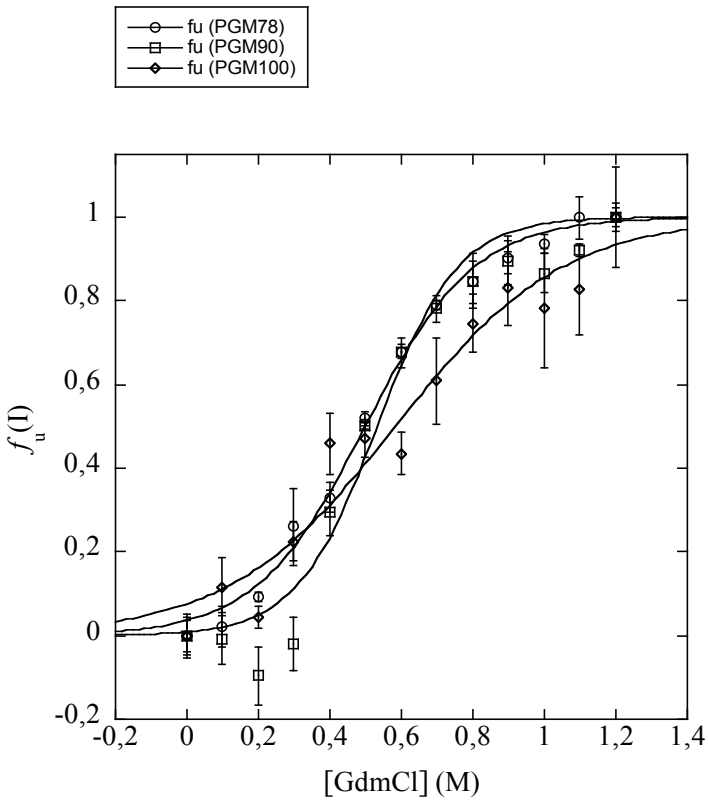
GTC CAT GAT TTT GGT GCT GCA TTT GAT GGC# GAC# GGC#
gtaagttaattagtttgtgatagatgtct
tattatttttatggtaaaactgacaagtatagctagaaaaatactttcaaaggatgtatagttc
tagaattgttgataaaaaacaattcatttatttcagttgcttatatatttgtttgtgtggctga
ggtgctgattgtctcacatctcaaaatgttttggtgtttggtgactttt **ag** GAC AGA AAC
ATG ATA CTT GGT AAG AAT GGC TTC TTT GTA TCG CCA TGT GAC TCC CTG
GCA GTC ATA GCT GCA CAT TTG GAG TGT ATA CCA TAT TTC AAG AAG TCT
GGC ATA AAA GGT TAT GCC **AGA** AGC ATG CCA ACT AGT GGG GCC ATT GAT
AG **gt**aaatatataaaaatgatcatctgtgg
acctaattctgttatgtttttcagttcttttgttttaccttacttcatctcatatatagtagtc
acagatgatatattgatcatcttattctgaactaggctacaaaataagagtttgccagtctaat
aaattaaaattatcacactaactgctgctttttatgatac**ag** **A** GTG GCA CAG AAG AAG
GGC AAG GAG ATG TTT GAG GTG CCA ACA GGC TGG AAG TTC TTT GGT AAC
CTT ACG GAT GCT GGT AGA CTT TCA CTT TGT GGA GAG GAG AGC TTT GGG
ACA GGA TCA GAT CAC ATC AG **gt**aatatt
acacaatgctacatgtgtcaaatttcattagaaataacttcatcttgtaagttgtatacacttt
ctgtaagccgactagagtttccatttcagaattcatttaacatgttatcagtcatatatattct
atatacggatttgaattttgtaatgtgagctatccatagctgatattgacatattgctgctcag
ttgcctatatgagactaaggcagtcttatctctcattcttggatgctgctatctataaaaccat
tttgacataatacatcatcaaagtgtgtgtttgtattacatcattatgtgttgaaatgttgaac
tgtttattgtgtttc**ag**A **GAA** AAG GAT GGC TTA TGG GCA GTT TTG GCC TGG
CTC TCT ATT TTA GCC TGC AAG AAG CAG TCA GTA GAA GAA ATC TTA AAG
GAT CAC TGG AAG ACA TAT GGC AGG AAC TTC TTC ACT AG
gttaatatcatagatgtattacagtctcataaaatc
atgtataaaatttataatatttaatcacagttttgctgtaataatatgaaaacctacttgaaat
tgataagatgtgaaaaataactttgtggctgttacttgtaccttctttatttgttgtggtgact
gacagttgcacctatttgc**ag** G TAT GAT TAT GAG AAT GTT GAA TCT GAT CCA
GCC AAT CAG ATG ATG GCC AAC CTT GAT AAA ATG GCA GCT GAC TCA TCT
ATT GTT GGC AAG GTG TTC AGT CAT GGT GAC AAA TCA TAT AAA GTA GCC
AAG ATG GAC AAC TTT GAA TAC ACT GAC CCA ATT GAC AAC AGT GTA TCA
AAA AAA CAG **gt**aattctgtcatct
tattaattgtgcacacatacacacctttgttcacatttgtgttaattatttgctctgtaaaata
aatactgcccaattcagaatgaaaaaaaaatgtctttcttacatcatgcaattgttattttgtg
tcacctatat**ag** GGC ATC CGG ATC ATT TTT GAG GAT GGA TCA AGG ATT ATA
TTC CGT CTG AGT GGT ACA GGA AGT GCT GGA GCA ACA ATC AGG ATG TAC
ATT GAT AGC TAT GAG TCA GAT TCA AAC AAA CAG CTC CTA GAT TCT CAG
gtttgttcacagcttaaatataacaagtgttatat
attctaaatttgcagcattgatgctcaatattgtgattaatattttagatataattttatatta
tgaaagaccttatcattaccactgcttatgaactgattgtttatgtattaaagtagtttcttga
gtggttattagggaaggcatgatgtcacttcattttcattctaatgttagcactatataaaatg
ctttaaaaatctatgcacaactcccactgaacaaatatttatttaattatcattccattctgc**a**
g GTC ATG CTG AAA CCA CTG ATT GAG ATA GCA CTG GAA ATA TCC CAG
CTT AGA GAG CTG ACA GGA AGA CAG CAG CCT ACT GTT ATT ACT TAA
atttggtatgcagccaacatttttgtcttcaatta
ccatgtagtgctatgtcatgtgatgagctatgcttagatgatctgtatgtaaaaaaaaaaaaaa
aaaaaaaaaaaaaaaaaaaaaaaaa

(2) translated sequence of the AP-PGM enzyme (562 aa). Bold uppercase letters between parentheses are the three polymorphic sites for which the alternative allele has a frequency greater than 5%. Bold lowercase letters correspond to alternative mutations, which frequency is lower than 5% in genotyped regions. Other bold sites represent non-synonymous singletons in our set of sequenced individuals and, grey zones correspond to intronic regions linking the translated regions. *: stop codon.

MSLKSV**T**VATK**P**FD**G**QKPGTSGLRKATKIF**M**Q**E**HYT**(E/q)**NF**(V/L)**QCTLSA**M**G**D**KLKGCT**L**VVGG**D**G
RYYGKEA**S**EKIIKMC**A**G**N**G-i1-VAKVIIGK**D**GILSTPAVSCLIR**K**NHTDGGIILTASHNPGGPNADFGI**K**
FNIANG-i2-GPAPAGVTDHIYALTQKI**T(E/Q)YH**TVPDLKADICT**I**GSQTFTVDDHPFNIEV**ID**S**V**
(E/Q)DY**MD**YMKEIFDFNSIRGLLTGEGGQTKLKV**L**VNALSGVVGPYVKRILCQELGMDEASAVNCVPLDF
GGGHPDPNL**TY**AADLVNELKK**G**VH**D**FGAAFDGDG-i3-DR**(N/d)MI**LGKNGF**F**VSPCDSLAV**I**AAHLE
CIPYFKKSGIKGYA**R**SMPTSGAID(**R/i**)-i4-VA**Q**KKGKE**M**FEVPT**(G/s,d)**WKFFGNL(**M/t**)DA**G**RL
S**L**CGEESFGTGS**D**HIR-i5**-(E/k)**KDGLWAVLAWLSIL**A**CKKQSVE**E**ILKDHWKTY**G**RNFF**T**R-i6-YDY
ENVESDPANQMMANLDK**M**AADSSIVGKVFSHGDKSYKVAKMDNFEYTDPIDNSVSKKQ-i7-GIR**(I/l)**IF
EDGSRI(**I/t**)**(F/l)**RLSGTGSAGATIRMYIDSYESDSNKQ**L**LDS-i8-QV**M**LKPLIEIALEISQLRELTG
RQQPTVIT*

FigS2. Guanidium chloride (GdmCl) denaturation curves for the three PGM1 overexpressed isoforms. The fluorescence intensity at 324 nm (excitation at 290 nm) is presented as a function of GdmCl concentration. (A) Evolution of the protein denaturation along the GdmCl gradient of isoform QE (78) showing the normal (N), intermediate (I) and denaturated (D) states of the protein. (B) Denaturation curves f_u *(I)* and f_u *(II)* of the three isoforms showing that isoform differences stand during the first step of protein denaturation.




fu (PGM78)
fu (PGM90)
fu (PGM100)
f_u(I)
[GdmCl] (M)

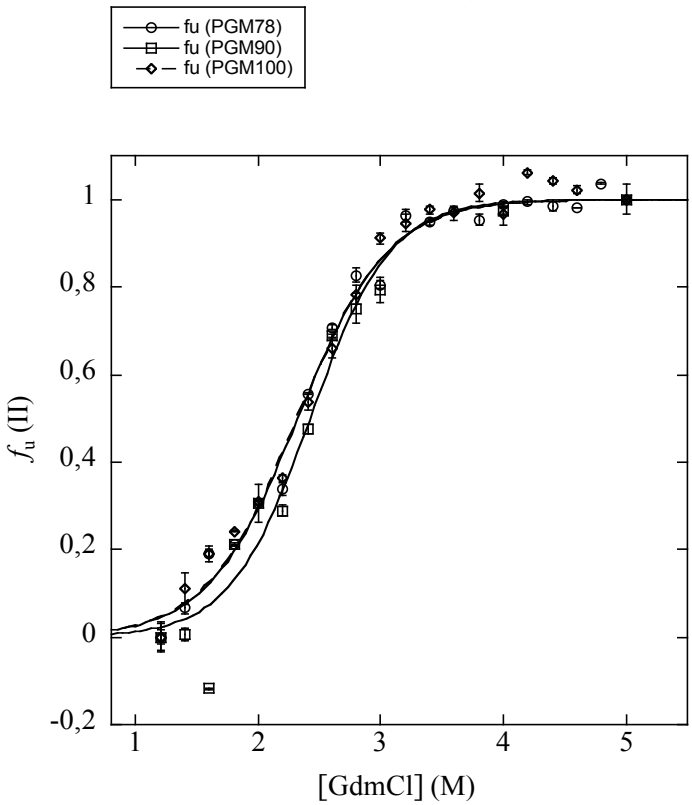

fu (PGM78)
fu (PGM90)
fu (PGM100)
f_u (II)
[GdmCl] (M)

Fig S3. Modelled regression curves of the folded/unfolded protein states f_u(I) and f_u (II) of three PGM1 overexpressed isoforms for each transition according to the denaturation equilibrium N↔I↔U (Native↔Intermediate↔Unfolded).

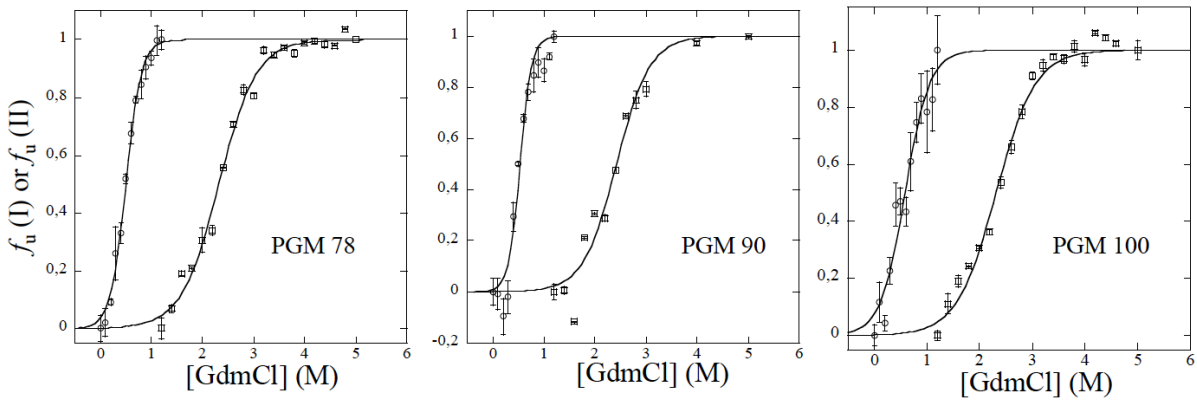
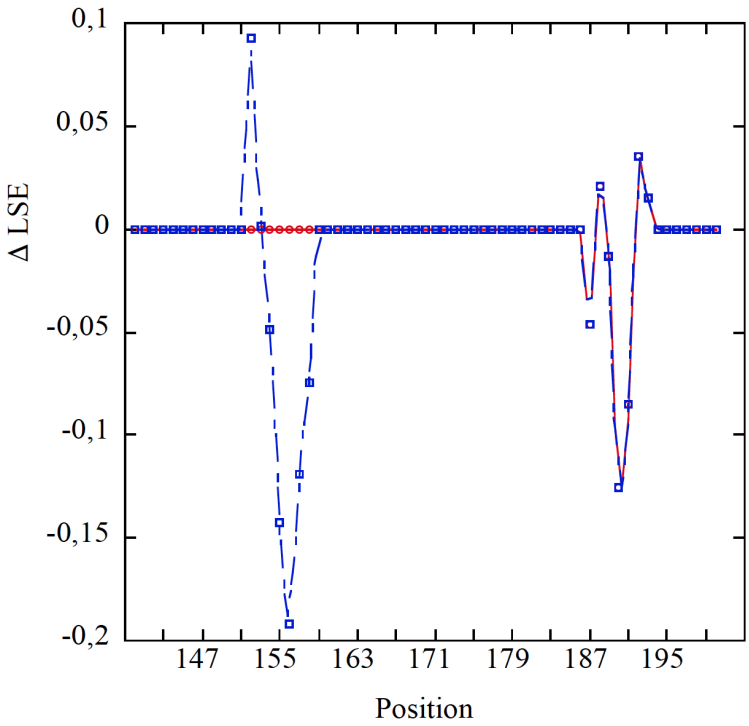



Fig S4. Effect of substitutions E155Q and E190Q on the local structural entropy of the molecule showing that the replacement of a glutamate to a glutamine has more impact on genotype 90. Calculation: ΔLES (O) = LSE PGM 90 – LSE PGM 100 ΔLES (□) = LSE PGM 90 – LES PGM 78 (http://SDSE.life.nctu.edu.tw).