Developmental & Comparative Immunology
April 2015, Volume 49, Issue 2, Pages 231-238
http://dx.doi.org/10.1016/j.dci.2014.12.002
http://archimer.ifremer.fr/doc/00239/35009/

High polymorphism in big defensin gene expression reveals presence-absence gene variability (PAV) in the oyster *Crassostrea gigas*

Rosa Rafael D. [1,] *, Alonso Pascal [1], Santini Adrien [1], Vergnes Agnes [1], Bachere Evelyne [1]

[1] Ifremer, CNRS, Université Montpellier 2, IRD, and Université Montpellier 1 (UMR5119 - Ecology of Coastal Marine Systems), Place Eugène Bataillon, CC80, 34095 Montpellier, France

* Corresponding author : Rafael D. Rosa, tel.: +55 48 3721 6163 ; Fax: +55 48 3721 5528 ; email address : rafael.d.rosa@ufsc.br

Abstract :

We report here the first evidence in an invertebrate, the oyster Crassostrea gigas, of a phenomenon of Presence-Absence Variation (PAV) affecting immune-related genes. We previously evidenced an extraordinary interindividual variability in the basal mRNA abundances of oyster immune genes including those coding for a family of antimicrobial peptides, the big defensins (Cg-BigDef). Cg-BigDef is a diverse family composed of three members: Cg-BigDef1 to -3. Here, we show that besides a high polymorphism in Cg-BigDef mRNA expression, not all individual oysters express simultaneously the three Cg-BigDefs. Moreover, in numerous individuals, no expression of Cg-BigDefs could be detected. Further investigation at the genomic level revealed that in individuals in which the transcription of one or all Cg-BigDefs was absent the corresponding Cg-bigdef gene was missing. In our experiments, no correlation was found between Cg-bigdef PAV and oyster capacity to survive Vibrio infections. The discovery of P-A immune genes in oysters leads to reconsider the role that plays the immune system in the individual adaptation to survive environmental, biotic and abiotic stresses.

Highlights

► *Cg*-BigDefs display a high interindividual polymorphism in basal mRNA expression. ► All three *Cg-BigDefs* are not simultaneously expressed in a single oyster. ► Lack of *Cg-BigDef* mRNA expression is likely due to the absence of the corresponding gene. ► First report in an invertebrate of Presence-Absence Variation affecting immune genes.

Keywords : mollusk bivalve, antimicrobial peptide, polymorphism, gene copy number variation, high-throughput microfluidic RT-qPCR, Vibrio infection

1. Introduction

Understanding the mechanisms of resistance or susceptibility to infections in the oyster *Crassostrea gigas* should contribute to identify the causes of mass mortalities that affect this marine mollusk species worldwide. In this context of multifactorial disease, significant progresses were made in characterizing the oyster immune system and responses to biotic and abiotic environment factors. Particular attention has been paid to the mechanisms involved in the capability of the oyster to overcome infections (de Lorgeril et al., 2011; Schmitt et al., 2012) but also to the genetic bases that govern oyster survival (Rosa et al., 2012).

High-throughput transcriptomic studies have revealed in *C. gigas* great individual polymorphism in the expression of many immune-related genes (Rosa et al., 2012). This supported previous evidence that *C. gigas* is a highly polymorphic and intron-rich species in terms of number of SNPs and repetitive DNA sequences (Kim et al., 2011; Saavedra and Bachère, 2006; Sauvage et al., 2007). Thus, besides sequence polymorphism, the oyster displays great polymorphism of expression that can be related to gene Copy Number Variation (CNV) in genome contents between different individuals (Rosa et al., 2012; Schmitt et al., 2013). CNV are known to be of fundamental importance for genetic variation in populations, evolution of genome complexity and evolution of genes with novel functions. Many studies have looked for associations between the copy number of genes and disease susceptibility (reviewed in Cantsilieris and White, 2013). For instance, in human, whereas extensive copy number of β-defensin antimicrobial peptide has been associated to the risk of psoriasis (Hollox et al., 2008), low copy number of β-defensin has been reported to predispose to Crohn disease (Fellermann et al., 2006).

Surprisingly, our high-throughput transcriptomic studies have also revealed numerous genes whose basal levels of expression could not be detected in some individuals (Rosa et al., 2012). Those genes were shown to be involved in immune responses such as genes encoding

members of the big defensin antimicrobial peptide family, *Cg*-BigDef (Rosa et al., 2011). Oyster big defensins (*Cg*-BigDef) have been shown to be a diverse antimicrobial peptide family composed of three members (*Cg*-BigDef1, *Cg*-BigDef2 and *Cg*-BigDef3) that are encoded by distinct genomic sequences (Rosa et al., 2011). An interesting particularity of the oyster *Cg*-BigDef family is that their members follow different patterns of gene regulation upon microbial challenge. Indeed, whereas *Cg-BigDef1* and *Cg-BigDef2* are up-regulated in circulating hemocytes in response to pathogenic *Vibrio* infection, *Cg-BigDef3* is constitutively expressed and non-regulated (Rosa et al., 2011). In this study, an unusual variability in mRNA expression was evidenced in oyster sampling for all *Cg-BigDefs*. As *Cg-BigDef1* and -*2* transcriptions are highly induced by microbial molecular patterns (Rosa et al., 2011), it was reasonable to expect that these molecules could be implicated in preventing from development infectious processes. This prompted us to further analyze such polymorphism of expression in *C. gigas*.

Here, a detailed analysis of *Cg-BigDef* mRNA expression at individual level confirmed firstly a high interindividual variability in basal gene expression for the three *Cg-BigDef* forms. In addition, we found that the three *Cg-BigDefs* are not simultaneously transcribed in a single oyster. More surprisingly, in some individual oysters, basal mRNA levels were not detected for any *Cg-BigDefs*. Besides, we evidenced that the absence of *Cg-BigDef* transcription is likely associated to the lack of *Cg-bigdef* genes, revealing the extreme form of CNV named Presence-Absence Variation (PAV). Finally, the Presence-Absence Variation of *Cg-bigdefs* in the oyster genome was shown to be not related to a better capacity of the oysters to survive experimental *Vibrio* infections.

2. Materials and Methods

2.1. Oysters and experimental *Vibrio* infections

Adult two-year old *Crassostrea gigas* oysters were purchased from a local oyster farm in Mèze (Gulf of Lion, France) and kept in aquaria containing filtered seawater at 15°C at the Aquaculture Experimental Platform of Ifremer, Palavas-Les-Flots. During two weeks, animals were fed twice with a mixture of live microalgae (*Nannochloropsis oculata*: 2.5×10^5 cells/ml; *Tetraselmis suecica*: 2.5×10^2 cells/ml). Following the acclimatation period, 90 oysters were individually tagged and hemolymph (~500 µl per animal) was withdrawn from the adductor muscle without causing the death of the animals. Hemocyte samples were individually collected by centrifugation (1,500 ×g, 15 min, 4°C), homogenized in 1 ml of TRIzol® reagent (Invitrogen) and frozen at -20°C until RNA extraction. After hemolymph collection, animals were placed in glass tanks (20 animals per tank) and allowed to recover for 8 days prior to experimental infections. Then, animals were intramuscularly injected with 5×10^8 CFU/animal of the virulent *V. tasmaniensis* LGP32 (=*V. splendidus* LGP32). Mortalities were monitored daily and individually tagged animals that did not survive were noted and discarded. After the end of acute mortalities (96 h post-infection), stored hemolymph samples were categorized according to oyster survival (surviving *versus* non-surviving). The standardization of the experimental infections and the preparation of the bacterial inoculum were performed as previously described (Duperthuy et al., 2010). All experimentations were performed according to the Ifremer animal care guideline and policy.

Hemocyte samples from 73 additional oysters (naive animals) were obtained as described above, homogenized in 1 ml of TRIzol® reagent (Invitrogen) and frozen at -20°C until RNA extraction. From those, 23 oysters were randomly chosen, removed from their shells and immediately frozen in liquid nitrogen. Frozen whole-body oysters were individually ground to fine powder for further extractions of total RNA and genomic DNA in same individuals.

121 **2.2. Genomic DNA isolation**

122 Genomic DNA (gDNA) was isolated individually as previously described (Schmitt et
123 al., 2010). Briefly, 50-100 mg of oyster powder were incubated overnight at 55°C in 500 µl of
124 DNA extraction buffer (100 mM NaCl, 10 mM Tris-HCl pH 8, 25 mM EDTA pH 8, 0.5%
125 SDS, 0.1 mg/ml protease K). Following a step of phenol/chloroform extraction, gDNA
126 samples were precipitated by addition of cold ethanol and treated with 50 µg/ml RNase A
127 (Invitrogen) for 30 min at 37°C to eliminate contaminating RNA. Quantification and quality
128 of gDNA samples were assessed by spectrophotometry (NanoDrop ND-1000 Thermo
129 Scientific) and 0.8% agarose gel electrophoresis, respectively.

130

131 **2.3. Total RNA extraction and cDNA synthesis**

132 Total RNA was extracted individually by using TRIzol® reagent (Invitrogen) according
133 to the manufacturer's instructions. RNA samples were then treated with DNase I (Invitrogen)
134 for 15 min at room temperature to eliminate contaminating genomic DNA. After DNase I
135 inactivation (10 min at 65°C), samples were precipitated with 0.3 M sodium acetate. RNA
136 amount and quality were assessed by spectrophotometric analysis and the integrity of total
137 RNA was analyzed by 0.8% agarose gel electrophoresis.

138 Following heat denaturation (70°C for 5 min), reverse transcription was performed
139 using 250 ng of purified total RNA with 50 ng/µl oligo$(dT)_{12\text{-}18}$ in a 20-µl reaction volume
140 containing 1 mM dNTPs, 1 unit of RNaseOUT Ribonuclease and 200 units M-MLV reverse
141 transcriptase in reverse transcriptase buffer according to the manufacturer's instructions
142 (Invitrogen). Attempting to increase the number of cDNA templates, a supplementary step of
143 preamplification was performed using 1.25 µl of cDNA, 1.25 µl of a mixture of forward and
144 reverse primers (200 nM each) and 2.5 µl of 2× TaqMan®PreAmpMaster Mix (Applied
145 Biosystems). The specific primer pairs for each *Cg*-BigDef form are listed in **Table 1**. PCR

cycling conditions were as follows: polymerase activation at 95°C for 10 min, followed by 14 cycles of 95°C for 15 s and 60°C for 4 min. Preamplification products were diluted to 1:5 in TE buffer (10 mM Tris, 1 mM EDTA, pH 7.5) and stored at -20°C.

2.4. High-throughput RT-qPCR using 96.96 microfluidic dynamic arrays

The relative gene expression was measured by using the high-throughput microfluidic RT-qPCR platform BioMark™ (Fluidigm) (Jang et al., 2011). The sample reaction mixtures were performed in a final volume of 5 µl containing 1.25 µl of preamplified cDNA (diluted 1:5), 2.5 µl of 2× TaqMan®Gene Expression Master Mix (Applied Biosystems), 0.25 µl of 20× DNA Binding Dye Sample Loading Reagent (Fluidigm), 0.25 µl of 20× EvaGreen (Biotium) and 0.75 µl of TE buffer. Primer reaction mixtures were made in the same volume of 5 µl containing 2.5 µl of 2× Assay Loading Reagent (Fluidigm), 1.25 µl of 20 µM of forward and reverse primer mix and 1.25 µl of TE buffer. Both sample and primer reaction mixtures were loaded into the dynamic array chip that was subsequently placed on the NanoFlex™ 4-IFC Controller for loading and mixing. After approximately 50 min, the chip was transferred to the BioMark™ Real-Time PCR System.

The cycling program used consisted of 10 min at 95°C followed by 40 cycles of 95°C for 15 s and 1 min at 60°C. Melting curves analysis was performed after completed RT-qPCR collecting fluorescence between 60-95°C at 0.5°C increments. Data were analyzed using the BioMark™ Real-time PCR analysis software to obtain Cq values. Results are presented as changes in relative expression normalized with the Cq geometric mean value of the eukaryotic translation elongation factor 1-alpha (*Cg-ef1α*: AB122066), the ribosomal protein L40 (*Cg-rpl40*: FP004478) and the 40S ribosomal protein S6 (*Cg-rps6*: CU686508) as endogenous reference genes (Livak and Schmittgen, 2001). Statistical significance between expression

detection of *Cg-BigDefs* and oyster capacity to survive *V. tasmaniensis* LGP32 infections was determined by ANOVA test at $p<0.05$.

2.5. Genotyping of *Cg-bigdefs*

PCR amplifications for *Cg-bigdef* genotyping were conducted in a 25-µl reaction volume using 1 µl of synthesized cDNA or 30 ng of purified gDNA as template. Primer sequences are listed in **Table 1**. PCR conditions were as follows: 10 min at 94°C, then 30 cycles at 94°C for 1 min, 56°C for 1 min, 72°C for 3 min and a final elongation step of 72°C for 10 min. PCR products were analyzed by electrophoresis on 1.5% agarose gel stained with ethidium bromide. The amplification products were cloned into a pCR®-Blunt II-TOPO® cloning vector using a Zero Blunt® TOPO® PCR cloning kit (Invitrogen). The positive recombinant clones were identified by colony PCR and were sequenced in both directions.

2.6. Sequence data analysis

Obtained sequences were analyzed for similarities using BLASTX at the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov). The multiple alignments were generated by using the ClustalW2 Multiple Alignment Program (ClustalW2; http://www.ebi.ac.uk/Tools/msa/clustalw2/). The phylogenetic analyses based on the full-length cDNA and genomic sequences of *C. gigas* big defensins were performed using the Neighbour-Joining method with the software MEGA version 4 (Tamura et al., 2007). Bootstrap sampling was reiterated 1,000 times.

3. Results

3.1. Oyster big defensins display a high interindividual polymorphism in basal gene expression

The basal expression levels of the three big defensins were first analyzed by a high-throughput RT-qPCR device in 163 individual oysters (**Fig. 1**; **Appendix A**). Specific primers were used for each *Cg-BigDef* form. The RT-qPCR analyses evidenced a high interindividual variability for the three *Cg-BigDef* transcripts. The *Cg-bigdef1* and *Cg-bigdef2* forms whose expression is inducible showed lower basal expression compared to the *Cg-bigdef3* form, constitutively expressed. Figure 2A shows the relative mRNA abundance of *Cg-BigDefs* in ten individual oysters which are representatives of individual differences in basal gene expression of the 163 oysters analyzed in this study. Depending on individuals, differences reached up to 30, 27 and 7-fold for *Cg-BigDef1*, *Cg-BigDef2* and *Cg-BigDef3*, respectively (**Fig. 1**; **Fig. 2A**). Besides, the three *Cg-BigDef* forms were never found to be simultaneously expressed among the 163 oysters analyzed (**Table 2**). Interestingly, only one or two forms of *Cg-BigDefs* were shown to be expressed in a same individual. *Cg-BigDef1* and *Cg-BigDef2* were seen in 81 and 47 individuals, respectively, while *Cg-BigDef3* was expressed in 21 oysters only. Similar results were obtained when using different specific primer pairs for each *Cg-BigDef* form (**Table 1**).

More surprisingly, among the 163 oysters, 39 did not express any *Cg-BigDefs* (**Table 2**). This lack of expression was further confirmed by RT-qPCR analyses using two universal primer pairs common to all *Cg*-BigDef sequences. As a control to these RT-qPCR analyses, the expression of three reference genes (*Cg-rpl40*, *Cg-ef1α and Cg-rps6*) was found to be uniform across all tested individuals (**Appendix A**). These results evidenced a high interindividual variability in gene expression of *C. gigas* big defensins, including cases where their basal expression is fully absent in some individuals.

3.2. Lack of *Cg-BigDef* expression is likely associated to the absence of encoding genes in the oyster genome

In light of the degree of polymorphism in gene expression observed for *Cg*-BigDefs, we further investigated the relationships between the expression and presence of the three *Cg-bigdef* genes in 23 individuals from which we have extracted both mRNA and gDNA. Detection of transcripts was performed by RT-qPCR and detection of corresponding genes by conventional PCR with specific primer pairs for each form (**Table 1**). The expression profiles obtained for these 23 individuals confirmed the high variability in basal expression levels observed above by high-throughput RT-qPCR analyses (**Fig. 2A**; **Appendix A**). Figure 2B shows the amplification of *Cg-bigdef* genes in ten individual oysters which are representatives of the 23 oysters analyzed. Interestingly, genes encoding a given *Cg*-BigDef were only amplified in individuals expressing the corresponding form at basal levels (**Fig. 2B**). The accuracy of these results was further confirmed by using an additional primer pair specific for the different *Cg-bigdefs* and two universal primer pairs (common to all forms). Besides, as a control of the integrity of extracted gDNA samples, *Cg-ef1α* gene was successfully amplified in all oysters analyzed (**Fig. 2B**).

3.3. *Cg*-BigDef forms are likely encoded by three distinct genes

We performed an exhaustive gDNA and cDNA cloning and sequencing of the *Cg-bigdefs* from four individuals expressing one or two forms, first to ensure the authenticity of the PCR amplifications but also to gain insights into the genomic organization of the big defensin family. We sampled one individual that transcribed simultaneously *Cg-BigDef1* and *-2* as well as one individual that was shown to transcribe both *Cg-BigDef1* and *-3*. Two additional individuals expressing only *Cg-BigDef2* or *-3* were also sampled. Both gene and cDNA sequences were obtained from these four individuals where the *Cg-BigDefs* were detected at the level of transcripts and genes. The obtained sequences were deposited in GenBank under the accession numbers JN251121 to JN251132.

Multiple alignments of the *Cg*-BigDef gDNA and cDNA sequences obtained in single individuals confirmed that each oyster big defensin form is transcribed by a separated gene. Indeed, whereas *Cg-bigdef1* and *Cg-bigdef2* genes included two exons interrupted by a single intron, an additional intron and exon were observed in *Cg-bigdef3* upstream the first exon common to the other *Cg-bigdef* genes (**Fig. 3**). Interestingly, in one individual in which both *Cg-BigDef1* and *-2* were expressed, two different *Cg-bigdef1* sequences were identified in addition to the *Cg-bigdef2* gene (**Fig. 4**). In this oyster, the *Cg-bigdef1* genes differed in the length of the intron sequence and by the presence of two nucleotide substitutions (**Fig. 4**). Moreover, for the other individuals analyzed, single nucleotide polymorphisms (SNP) were found between transcripts and genes, which was evocative of two different alleles of a same *Cg-bigdef* gene. Thus, each *Cg-bigdef* would correspond rather to a distinct gene than to alleles from a same *Cg-bigdef* locus.

Sequence analysis revealed that, for each *Cg-bigdef* gene, the length of exons was well conserved among individuals whereas the introns lengths were variable. According to individuals, SNP were identified in both intron and exon sequences for a same *Cg-bigdef* gene. Additionally, introns from all *Cg-bigdef* genes included some microsatellite sequences (such as CTAT, CT and CA) and the canonical GT/AG splice site junctions (**Fig. 3**). The intron sequences of *Cg-bigdef3* appeared to differ considerably in both length and nucleotide composition from those of *Cg-bigdef1* and *Cg-bigdef2* genes (**Fig. 3**).

A phylogenetic tree was built with the nucleotide sequence of six *Cg-bigdef* gene sequences obtained here and with *Cg-bigdef* gene sequences previously deposited in GenBank (Rosa et al., 2011). In this tree, *Cg-bigdef1* and *Cg-bigdef2* genes clustered in a same clade distinct from the *Cg-bigdef3* gene (**Fig. 5**). An additional phylogenetic tree constructed from cDNA sequence data confirmed that *Cg*-BigDef1 and *Cg*-BigDef2 forms are more related to each other than to *Cg*-BigDef3 (**Fig. 5**). Altogether, results from sequencing and phylogenetic

analyses strongly suggest that *Cg*-BigDefs constitute a multigenic family of antimicrobial peptides in *C. gigas* oysters.

3.4. Presence-absence variation (PAV) in *Cg-BigDef* gene expression is not associated to oyster capacity to survive virulent *Vibrio* infections

In order to investigate whether the presence or absence of *Cg*-BigDefs was associated to an inherent capacity of oysters to survive infections, we have quantified the mRNA levels of all three *Cg-BigDefs* (*Cg-BigDef1*, *Cg-BigDef2* and *Cg-BigDef3*) in 90 individual animals before a challenge with an oyster pathogen. By using a nonlethal sampling method prior to the experimental infections (Rosa et al., 2012), we have assessed the basal transcription profile of all three *Cg-BigDefs* in oysters able (S) and not able (NS) to survive infections by the virulent *V. tasmaniensis* LGP32 (=*V. splendidus* LGP32). Three independent experimental infections were conducted and hemolymph samples from 45 oysters of each survival capacity phenotype were obtained.

The detection of *Cg-BigDef* transcripts and their mRNA abundance were compared in individual oysters from both S and NS phenotypes by high-throughput microfluidic RT-qPCR analyses. Transcripts of *Cg-BigDef1* were detected in 25 individuals from both S and NS phenotypes whereas transcripts of *Cg-BigDef2* were detected in 10 and 12 individuals from S and NS phenotypes, respectively (**Table 3**). *Cg-BigDef3* transcripts were detected in 4 individuals from the S phenotype and in 2 from the NS phenotype. For the S phenotype, 27 and 6 individuals expressed, respectively, only one or two *Cg-BigDefs* (*Cg-BigDef1* and *Cg-BigDef2* or *Cg-BigDef1* and *Cg-BigDef3*) and 12 oysters did not express any *Cg-BigDefs* (**Table 3**). For the NS phenotype, 4 and 31 oysters expressed, respectively, only one or two *Cg-BigDefs* (*Cg-BigDef1* and *Cg-BigDef2* or *Cg-BigDef1* and *Cg-BigDef3*) and 10 oysters did not express any *Cg-BigDefs*. No oysters expressing simultaneously *Cg-BigDef2* and *Cg-*

295 *BigDef3* or all three *Cg-BigDefs* were found in both S and NS phenotypes (**Table 3**). No
296 significant differences were found in the detection of *Cg-BigDefs* according to the S and NS
297 phenotypes, namely according to the oyster capability to survive virulent *Vibrio* infections
298 (p>0.05). Additionally, no significant differences in basal gene expression levels of the three
299 *Cg-BigDefs* were found between individuals from both survival capacity phenotypes.
300

301 **4. Discussion**

302 Taking advantages of the development of high-throughput genomic technologies, we
303 have explored the genetic determinants for *Crassostrea gigas* capacity to survive infections at
304 the individual level. These analyses highlighted an extraordinary interindividual
305 polymorphism in the basal expression of genes in this species (Rosa et al., 2012). Indeed,
306 most of the transcriptome profilings are performed with oyster sample pools, which hides
307 great information about oyster genomic diversity and complexity. Here, high-throughput RT-
308 qPCR analyses at the individual level revealed a high variability in basal gene expression
309 regulation of antimicrobial peptides belonging to *Cg*-BigDef family. Basal mRNA amounts of
310 both the inducibly expressed *Cg-BigDef1* and *Cg-BigDef2* and the constitutively-expressed
311 *Cg-BigDef3* forms appeared to be highly variable among individual oysters. For instance,
312 variations up to 30-fold were found in *Cg-BigDef1* expression according to individuals. This
313 interindividual polymorphism in *Cg-BigDef* transcription levels may reflect genetic structure
314 variation in gene copy numbers (CNV) as recently reported for other oyster AMPs, *Cg*-Defs
315 and *Cg*-Prps (Schmitt et al., 2013).

316 When analyzing the heterogeneity of transcription profiles of the big defensins, we
317 showed that no expression was detected in some oysters and that the lack of basal expression
318 was likely associated to the absence of the corresponding encoding sequences at the genome
319 level. Such results highlight structural variations in oyster genomes and particularly gene

Presence-Absence variation (PAV), a phenomenon described as the extreme case of Copy Number Variation (CNV) in which a large DNA segment present in some individuals can be entirely missing in others. To our knowledge, this is the first evidence of PAV in oysters.

Considering that *C. gigas* oyster displays a high degree of DNA sequence polymorphism and heterozygosity (Sauvage et al., 2007; Schmitt et al., 2010; Taris et al., 2008), a rigorous accuracy has been required in the choice of the primers used in this study. Two primer pairs specific for each of the three *Cg*-BigDefs have been designed as well as two primer pairs common for all three forms. The relationship between the presence-absence (P-A) of transcripts and the presence-absence of corresponding encoding gene has been confirmed in 23 individuals by different PCR analyses using eight different primer pairs spread across the entire sequences. Besides, the correspondence between the different genes and their corresponding transcripts amplified has been assessed by nucleotide sequencing. Thus, the non-amplification of any *Cg*-BigDef is likely due to the lack of genes instead of the existence of null alleles commonly reported in oysters (Reece et al., 2004).

Here, we have improved also knowledge about oyster big defensin diversity by the identification of novel variants for each *Cg*-BigDef form. Interestingly, in single individuals, we evidenced the presence of different variants for a same *Cg-bigdef* gene, suggesting the presence of different isoforms of a given *Cg*-BigDef in a same oyster. Likewise, the human *DEF* locus has been shown to be highly variable among individuals (Taudien et al., 2004). Variations in the human *DEF* locus include the presence of different alleles for a same *DEF* gene, and both CNV and PAV phenomena.

PAV appeared to concern all *Cg-bigdefs* but more particularly the constitutively expressed *Cg-bigdef3* gene that was recorded in less than 13% of the analyzed oyster population. Importantly, since we sampled animals from a same oyster hatchery, we cannot presume that the obtained frequency distribution of *Cg-bigdef* genes is representative of

natural oyster populations or representative of the species. The three *Cg-BigDefs* were never seen simultaneously expressed in a single oyster, probably due to the relatively limited oyster sampling size. Indeed, based on the occurrence frequency of the three *Cg-bigdefs* in this oyster population, the probability of one individual carrying the three loci was calculated to be less than 2%. Any of the three big defensin forms were detected in 23.9% (39/163) of the individuals examined. A single oyster can express only one or up two *Cg-BigDefs* concomitantly. Conversely, in *Litopenaeus* shrimp species all three members of the penaeidin family (*Litvan* PEN2, *Litvan* PEN3 and *Litvan* PEN4) and all four members of the anti-lipopolysaccharide factor (ALF) family (*Litvan* ALF-A, *Litvan* ALF-B, *Litvan* ALF-C and *Litvan* ALF-D) are simultaneously transcribed in a single individual (Cuthbertson et al., 2002; O'Leary and Gross, 2006; Rosa et al., 2013). Each penaeidin member is encoded by a separated gene and the interindividual molecular diversity is mainly generated by DNA polymorphism within each *PEN* gene locus (O'Leary and Gross, 2006). The simultaneous expression of different members of a given antimicrobial peptide family in a single animal has been shown to improve its antimicrobial defenses (Cuthbertson et al., 2002; Rosa et al., 2013).

Analyses of the *Cg-bigdef* gene sequences obtained from a single individual suggest that each *Cg*-BigDef form is encoded by a separate gene and not by different alleles from a same *Cg-bigdef* locus. Interestingly, the intron sequences of *Cg-bigdef3* were shown to be different from those of *Cg-bigdef1* and *Cg-bigdef2* genes, suggesting an ancient divergence between the constitutively and the inducibly expressed forms. Likewise to *Cg*-BigDefs, *Cg*-Defs and *Cg*-Prps are also multigenic families whose the different members are encoded by separate genes displaying distinct gene structures (Schmitt et al., 2010). Thus far, no transcription profiling of *Cg*-Defs and *Cg*-PRPs have been performed at the individual level but it can be assumed that PAV can also affect these effectors in oysters.

To date, intraspecific variations in gene contents such as PAV have never been reported in oysters and any other non-model invertebrate species, probably because few genome sequences are available. Indeed, whole *C. gigas* genome has been recently sequenced from one individual (Zhang et al., 2012). It is noteworthy that we failed to evidence in the published database any full sequences of big defensin encoding genes (Zhang et al., 2012). The PAV phenomenon for big defensins or any other genes in *C. gigas* deserves to be fully characterized by genetic mapping or through population-scale genome sequencing.

PAV is essentially known in human and particularly in plants in which this genetic diversity is exploited in domestication programs. In human, PAV has been described for the *DEFA3* gene encoding human neutrophil AMP HNP-3 (Linzmeier and Ganz, 2005). Interestingly, the occurrence of *DEFA3* gene is variable depending on the human geographical populations, and it has been suspected to be implicated with differences in infectious disease susceptibility (Ballana et al., 2007). In our study, no correlation has been shown between the P-A prevalence of the big defensins and the intrinsic capacity of the oysters to survive experimental *Vibrio* infections. However, one can expect that P-A genes others than those genes encoding AMP could have substantial involvement on oyster fitness and health in the context of multifactorial disease that currently affects this species worldwide.

In plants, regions of the genome encompassing PAV have been shown to contain gene clusters involved in disease resistance (Zhang et al., 2014). Thus, it is assumed that PAV may play important role in generating new pathogen resistances and they are now considered as polymorphic markers for genetic studies and breeding (Wang et al., 2014; Zhang et al., 2014). Overall, the great number of P-A genes observed in plants is recognized to play important roles in individual adaptation to external pressure and environmental stresses (Haun et al., 2011; Springer et al., 2009; Swanson-Wagner et al., 2010). Indeed, such structural variations

imply that the missing genes in some individuals are not necessary for survival but rather contribute to the genome evolution and to the interplay with the environment (Marroni et al., 2014). Undoubtedly, the first discovery of PAV that affect immune-related genes in oyster open questions about the functional consequences of such genome structural variation on the oyster immune responses but also on the individual capability to survive environmental, biotic and abiotic stresses.

Acknowledgments

We are in debt with Frédérique Cerqueira, Jean-Luc Rolland and Marc Leroy for their excellent technical assistance. We are also grateful to François Bonhomme for helpful comments and Delphine Destoumieux-Garzón for critical reading of the manuscript. This study received financial support from Ifremer and CNRS. RD Rosa was supported by a doctoral founding from CNPq-Brazil. Data used in this work were partly produced through molecular genetic analysis technical facilities of the SFR "Montpellier Environnement Biodiversité" and the "Platform qPHD UM2/Montpellier GenomiX".

References

Ballana, E., González, J.R., Bosch, N., Estivill, X., 2007. Inter-population variability of DEFA3 gene absence: correlation with haplotype structure and population variability. BMC Genomics 8, 14.

Cantsilieris, S., White, S.J., 2013. Correlating multiallelic copy number polymorphisms with disease susceptibility. Hum. Mutat. 34, 1–13.

Cuthbertson, B.J., Shepard, E.F., Chapman, R.W., Gross, P.S., 2002. Diversity of the penaeidin antimicrobial peptides in two shrimp species. Immunogenetics 54, 442–5.

de Lorgeril, J., Zenagui, R., Rosa, R.D., Piquemal, D., Bachère, E., 2011. Whole transcriptome profiling of successful immune response to *Vibrio* infections in the oyster *Crassostrea gigas* by digital gene expression analysis. PLoS One 6, e23142.

Duperthuy, M., Binesse, J., Le Roux, F., Romestand, B., Caro, A., Got, P., Givaudan, A., Mazel, D., Bachère, E., Destoumieux-Garzón, D., 2010. The major outer membrane protein OmpU of *Vibrio splendidus* contributes to host antimicrobial peptide resistance and is required for virulence in the oyster *Crassostrea gigas*. Environ. Microbiol. 12, 951–63.

Fellermann, K., Stange, D.E., Schaeffeler, E., Schmalzl, H., Wehkamp, J., Bevins, C.L., Reinisch, W., Teml, A., Schwab, M., Lichter, P., Radlwimmer, B., Stange, E.F., 2006. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. Am. J. Hum. Genet. 79, 439-48.

Haun, W.J., Hyten, D.L., Xu, W.W., Gerhardt, D.J., Albert, T.J., Richmond, T., Jeddeloh, J.A., Jia, G., Springer, N.M., Vance, C.P., Stupar, R.M., 2011. The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. Plant Physiol. 155, 645–55.

Hollox, E.J., Huffmeier, U., Zeeuwen, P.L., Palla, R., Lascorz, J., Rodijk-Olthuis, D., van de Kerkhof, P.C., Traupe, H., de Jongh, G., den Heijer, M., Reis, A., Armour, J.A., Schalkwijk, J., 2008. Psoriasis is associated with increased beta-defensin genomic copy number. Nat Genet. 40, 23–5.

Jang, J.S., Simon, V.A., Feddersen, R.M., Rakhshan, F., Schultz, D.A., Zschunke, M.A., Lingle, W.L., Kolbert, C.P., Jen, J., 2011. Quantitative miRNA expression analysis using fluidigm microfluidics dynamic arrays. BMC Genomics 12, 144.

441 Kim, W.-J., Jung, H., Gaffney, P.M., 2011. Development of type I genetic markers from
442 expressed sequence tags in highly polymorphic species. Mar. Biotechnol. (NY). 13, 127–
443 32.

444 Linzmeier, R.M., Ganz, T., 2005. Human defensin gene copy number polymorphisms:
445 comprehensive analysis of independent variation in alpha- and beta-defensin regions at
446 8p22-p23. Genomics 86, 423–30.

447 Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time
448 quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–8.

449 Marroni, F., Pinosio, S., Morgante, M., 2014. Structural variation and genome complexity: is
450 dispensable really dispensable? Curr. Opin. Plant Biol. 18, 31–6.

451 O'Leary, N.A., Gross, P.S., 2006. Genomic structure and transcriptional regulation of the
452 penaeidin gene family from *Litopenaeus vannamei*. Gene 371, 75–83.

453 Reece, K.S., Ribeiro, W.L., Gaffney, P.M., Carnegie, R.B., Allen, S.K., 2004. Microsatellite
454 marker development and analysis in the eastern oyster (*Crassostrea virginica*):
455 confirmation of null alleles and non-Mendelian segregation ratios. J. Hered. 95, 346–52.

456 Rosa, R.D., de Lorgeril, J., Tailliez, P., Bruno, R., Piquemal, D., Bachère, E., 2012. A
457 hemocyte gene expression signature correlated with predictive capacity of oysters to
458 survive *Vibrio* infections. BMC Genomics 13, 252.

459 Rosa, R.D., Santini, A., Fievet, J., Bulet, P., Destoumieux-Garzón, D., Bachère, E., 2011. Big
460 defensins, a diverse family of antimicrobial peptides that follows different patterns of
461 expression in hemocytes of the oyster *Crassostrea gigas*. PLoS One 6, e25594.

462 Rosa, R.D., Vergnes, A., de Lorgeril, J., Goncalves, P., Perazzolo, L.M., Sauné, L.,
463 Romestand, B., Fievet, J., Gueguen, Y., Bachère, E., Destoumieux-Garzón, D., 2013.
464 Functional divergence in shrimp anti-lipopolysaccharide factors (ALFs): from
465 recognition of cell wall components to antimicrobial activity. PLoS One 8, e67937.

466 Saavedra, C., Bachère, E., 2006. Bivalve genomics. Aquaculture 256, 1–14.

467 Sauvage, C., Bierne, N., Lapègue, S., Boudry, P., 2007. Single Nucleotide polymorphisms
468 and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. Gene
469 406, 13–22.

470 Schmitt, P., Gueguen, Y., Desmarais, E., Bachère, E., de Lorgeril, J., 2010. Molecular
471 diversity of antimicrobial effectors in the oyster *Crassostrea gigas*. BMC Evol. Biol. 10,
472 23.

473 Schmitt, P., Rosa, R.D., Duperthuy, M., de Lorgeril, J., Bachère, E., Destoumieux-Garzón, D.,
474 2012. The antimicrobial defense of the Pacific oyster, *Crassostrea gigas*. How diversity
475 may compensate for scarcity in the regulation of resident/pathogenic microflora. Front.
476 Microbiol. 3, 160.

477 Schmitt, P., Santini, A., Vergnes, A., Degremont, L., de Lorgeril, J., 2013. Sequence
478 polymorphism and expression variability of *Crassostrea gigas* immune related genes
479 discriminate two oyster lines contrasted in term of resistance to summer mortalities.
480 PLoS One 8, e75900.

481 Springer, N.M., Ying, K., Fu, Y., Ji, T., Yeh, C.-T., Jia, Y., Wu, W., Richmond, T., Kitzman,
482 J., Rosenbaum, H., Iniguez, A.L., Barbazuk, W.B., Jeddeloh, J.A., Nettleton, D.,

Schnable, P.S., 2009. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet. 5, e1000734.

Swanson-Wagner, R.A., Eichten, S.R., Kumari, S., Tiffin, P., Stein, J.C., Ware, D., Springer, N.M., 2010. Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res. 20, 1689–99.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–9.

Taris, N., Lang, R.P., Camara, M.D., 2008. Sequence polymorphism can produce serious artefacts in real-time PCR assays: hard lessons from Pacific oysters. BMC Genomics 9, 234.

Taudien, S., Galgoczy, P., Huse, K., Reichwald, K., Schilhabel, M., Szafranski, K., Shimizu, A., Asakawa, S., Frankish, A., Loncarevic, I.F., Shimizu, N., Siddiqui, R., Platzer, M., 2004. Polymorphic segmental duplications at 8p23.1 challenge the determination of individual defensin gene repertoires and the assembly of a contiguous human reference sequence. BMC Genomics 5, 92.

Wang, Y., Lu, J., Chen, S., Shu, L., Palmer, R.G., Xing, G., Li, Y., Yang, S., Yu, D., Zhao, T., Gai, J., 2014. Exploration of presence/absence variation and corresponding polymorphic markers in soybean genome. J. Integr. Plant Biol. 56, 1009–19.

Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., Xiong, Z., Que, H., Xie, Y., Holland, P.W.H., Paps, J., Zhu, Y., Wu, F., Chen, Y., Wang, J., Peng, C., Meng, J., Yang, L., Liu, J., Wen, B., Zhang, N., Huang, Z., Zhu, Q., Feng, Y., Mount, A., Hedgecock, D., Xu, Z., Liu, Y., Domazet-Lošo, T., Du, Y., Sun, X.,

505 Zhang, S., Liu, B., Cheng, P., Jiang, X., Li, J., Fan, D., Wang, W., Fu, W., Wang, T.,
506 Wang, B., Zhang, J., Peng, Z., Li, Y., Li, N., Wang, J., Chen, M., He, Y., Tan, F., Song,
507 X., Zheng, Q., Huang, R., Yang, H., Du, X., Chen, L., Yang, M., Gaffney, P.M., Wang,
508 S., Luo, L., She, Z., Ming, Y., Huang, W., Zhang, S., Huang, B., Zhang, Y., Qu, T., Ni,
509 P., Miao, G., Wang, J., Wang, Q., Steinberg, C.E.W., Wang, H., Li, N., Qian, L., Zhang,
510 G., Li, Y., Yang, H., Liu, X., Wang, J., Yin, Y., Wang, J., 2012. The oyster genome
511 reveals stress adaptation and complexity of shell formation. Nature 490, 49–54.

512 Zhang, L.-M., Luo, H., Liu, Z.-Q., Zhao, Y., Luo, J.-C., Hao, D.-Y., Jing, H.-C., 2014.
513 Genome-wide patterns of large-size presence/absence variants in sorghum. J. Integr.
514 Plant Biol. 56, 24–37.

515

516
517 **Figure captions**

518 **Figure 1**. Polymorphism in *Cg-BigDef* basal mRNA expression in 163 individual oysters.
519 Relative mRNA levels were analyzed in oysters expressing *Cg-BigDef1* (81/163), *Cg-BigDef2*
520 (47/163) and *Cg-BigDef3* (21/163). Relative expression data were normalized with the Cq
521 geometric mean value of three reference genes (*Cg-ef1α*, *Cg-rpl40* and *Cg-rps6*) according to
522 the $2^{-\Delta\Delta Cq}$ method.

523

524 **Figure 2**. The lack of detection of *Cg-BigDef* expression is likely due to the lack of the
525 corresponding gene in the oyster genome. (A) Relative mRNA abundance of the three *Cg-*
526 *BigDefs* in 10 individual oysters by RT-qPCR. Relative mRNA expressions were normalized
527 with the Cq geometric mean value of three reference genes (*Cg-ef1α*, *Cg-rpl40* and *Cg-rps6*)
528 according to the $2^{-\Delta\Delta Cq}$ method. (B) Detection of *Cg-bigdef* genes by conventional PCR using

specific (*Cg-bigdef1*, *Cg-bigdef2*, *Cg-bigdef3*) and universal (*Cg-bigdefs*) primer pairs. The quality and integrity of the gDNA samples were assessed by the amplification of the *Cg-ef1α* gene.

Figure 3. Each *Cg*-BigDef form is encoded by a separated gene. (A) A not-to-scale representation of *Cg-bigdef* genes. Black lines and boxes represent introns and exons, respectively. E=exon. (B) Multiple nucleotide alignment of the three *Cg-bigdef* genes. Exon sequences are in bold. The canonical GT/AG splicing recognition sites are shadowed with black backgrounds and the microsatellite sequences (CTAT, CT and CA) are indicated by rectangles.

Figure 4. Nucleotide alignment of the first exon and intron sequences from alleles of *Cg-bigdef1* and *Cg-bigdef2* genes from a same individual oyster. Differences (nucleotide substitutions or gaps) between the two *Cg-bigdef1* alleles and between *Cg-bigdef1* and *Cg-bigdef2* genes are shadowed with black and grey backgrounds, respectively. Exon sequences are in bold and the canonical GT/AG splice site junctions are indicated by rectangles.

Figure 5. Phylogeny of the oyster big defensin family. Phylogenetic trees were constructed based on (A) genomic and (B) cDNA sequences using the Neighbour-Joining method in MEGA 4. Bootstrap sampling was reiterated 1,000 times. Sequences included in these analyses were the following: (i) *Cg-bigdef* genes (GenBank: JF703155-JF703160, JN251122, JN251125, JN251126, JN251128, JN251131 and JN251132) and *Cg-defm* gene (GenBank: AM050547) as outgroup; (ii) *Cg-BigDef* transcripts (GenBank: JF703137-JF703154, JN251121, JN251123, JN251124, JN251127, JN251129 and JN251130) and *Cg*-Defh2 (GenBank: DQ400102) as outgroup.

554

555 **Supplementary data**

556 **Appendix A**. Cq values of three reference genes (*Cg-ef1α*, *Cg-rpl40* and *Cg-rps6*) and *Cg-*
557 BigDefs (*Cg-BigDef1*, *Cg-BigDef2* and *Cg-BigDef3*) analyzed by a high-throughput RT-
558 qPCR device in 163 individual oysters.

559 **Tables**

560 **Table 1**. Nucleotide sequence of primers.

Primer name	Forward primer (5'-3')	Reverse primer (5'-3')
Primers for RT-qPCR and gDNA amplification		
Cg-ef1α	GAGCGTGAACGTGGTATCAC	ACAGCACAGTCAGCCTGTGA
Cg-rps6	CAGAAGTGCCAGCTGACAGTC	AGAAGCAATCTCACACGGAC
Cg-rpl40	AATCTTGCACCGTCATGCAG	AATCAATCTCTGCTGATCTGG
Cg-bigdef1	TTCGCCTGCTTCCATACTGG	GTCATGGTCACTCCTTATTC
	TCTTTGCTGCTTTGGTGACC	GTCATGGTCACTCCTTATTC
Cg-bigdef2	TCTTTGCTGCTTTGGTGACC	CATAGTTTATCCCCTCCGTC
	TTCGCCTGCTTCCATACTGG	AATGACTGTCATGGTCAGAA
	AGAAGAAGGTGAGACGAG	TGATCCGCACACACCAAACG
Cg-bigdef3	AGAAGAAGGTGAGACGAG	GGCTGATTAATCCATGCAAG C
Cg-bigdefs	CAGTCTTTGCTGCTTTGGTG	TTGGCACACGAATGACTGTC
(universal)	TTCGCCTGCTTCCATACTGG	TTGGCACACGAATGACTGTC
Primers for molecular cloning and sequencing		
Cg-bigdef1	ACGTATAGGACTATCATGGAG	GTCATGGTCACTCCTTATTC
	TTCCTACCAAAGAATAAGGAG	CGTGGCTCAGTAATCTCTGC
	TGTTAACGTATAGGACTATC	AATGACTGTCATGGTCAGAA
Cg-bigdef2	TCTTTGCTGCTTTGGTGACC	CGTGGCTCAGTAATCTCTGC
	TGTTAACGTATAGGACTATC	CGTGGCTCAGTAATCTCTGC
Cg-bigdef3	GGAGAACTGTTAGTGTGCTG	GGCTGATTAATCCATGCAAG C

561

562

563

564 **Table 2**. Occurrence of basal *Cg-BigDef* expression in the oyster population analyzed.

Cg-BigDef	**Occurrence (%)**	
Cg-BigDef1 alone	60	(36.8%)
Cg-BigDef2 alone	31	(19%)
Cg-BigDef3 alone	08	(4.9%)
Cg-BigDef1 and *Cg-BigDef2*	12	(7.4%)
Cg-BigDef1 and *Cg-BigDef3*	09	(5.5%)
Cg-BigDef2 and *Cg-BigDef3*	04	(2.5%)
All three *Cg-BigDef* forms	0	(0%)
Absence of *Cg-BigDefs*	39	(23.9%)
Total of individuals	**163**	

565

566

567 **Table 3.** Occurrence of basal gene expression of *Cg-BigDefs* according to oyster survival

568 (surviving *versus* non-surviving) to virulent *Vibrio* infections

Cg-BigDef	**Occurrence (%)**			
	Surviving		Non-surviving	
Cg-BigDef1 alone	25	(55.5%)	25	(55.5%)
Cg-BigDef2 alone	10	(22.2%)	12	(26.6%)
Cg-BigDef3 alone	04	(8.8%)	02	(4.4%)
Cg-BigDef1 and *Cg-BigDef2*	27	(60.0%)	04	(8.8%)
Cg-BigDef1 and *Cg-BigDef3*	06	(13.3%)	31	(68.8%)
Cg-BigDef2 and *Cg-BigDef3*	0	(0%)	0	(0%)
All three *Cg-BigDef* forms	0	(0%)	0	(0%)
Absence of *Cg-BigDefs*	12	(26.6%)	10	(22.2%)
Total of individuals	**45**		**45**	

569

570

571

Figure 1

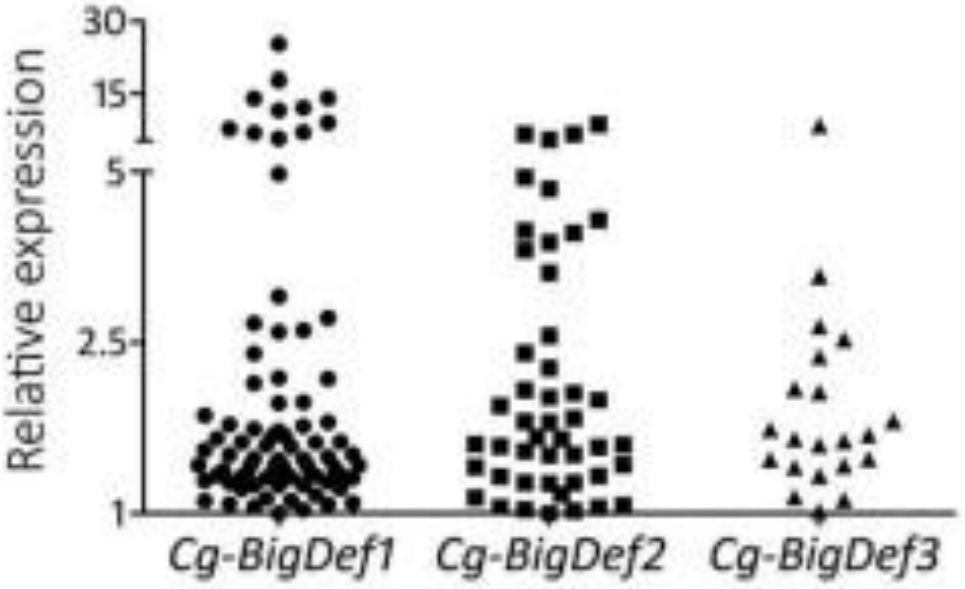

Figure 2

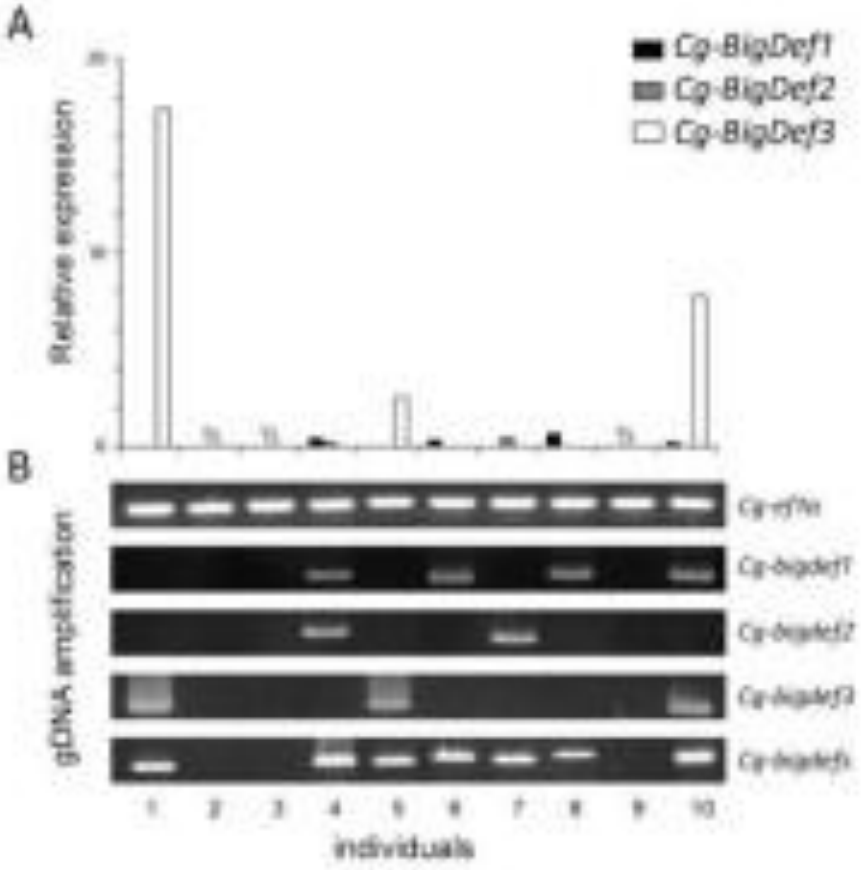

Figure 3

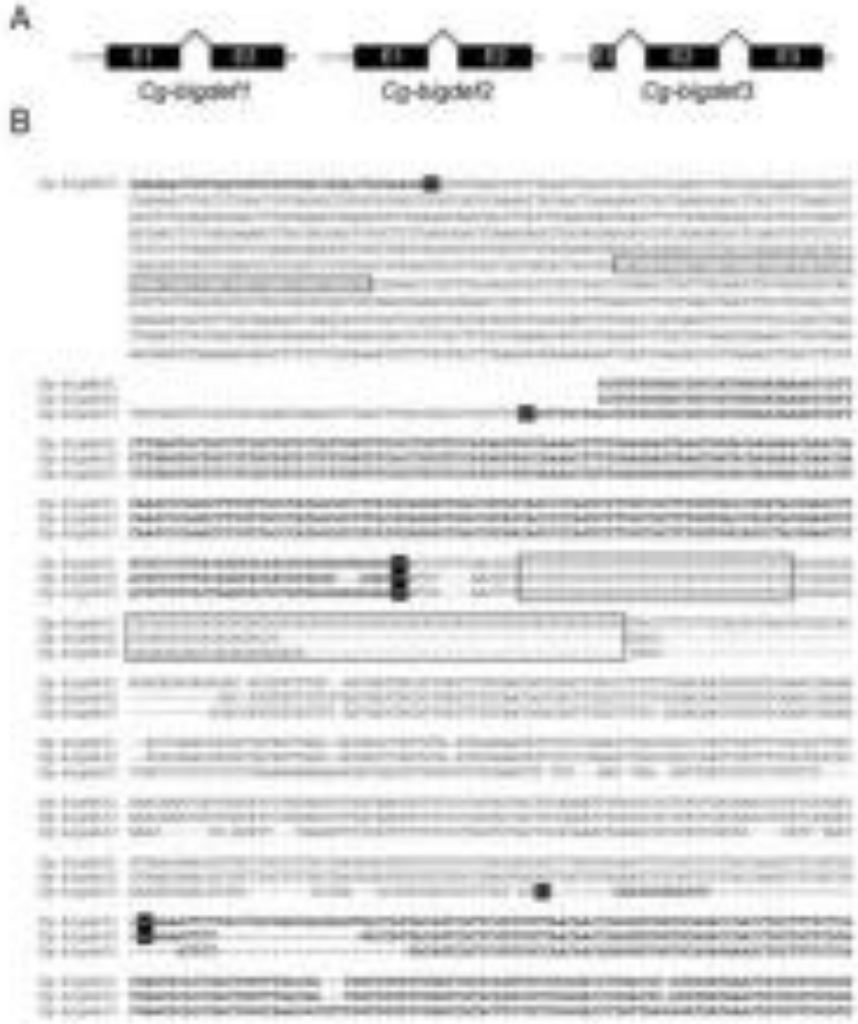

Figure 4

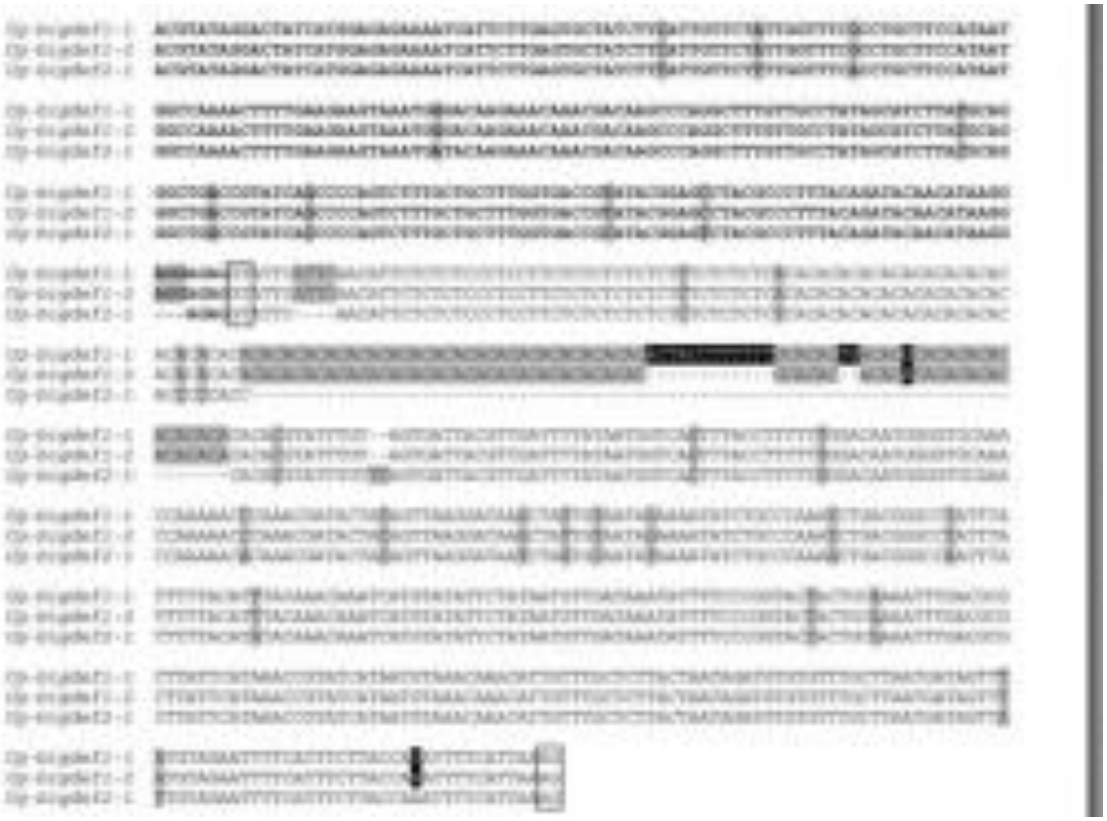

Figure 5