Marine Genomics 27 (2016) 17–23

Contents lists available at ScienceDirect

Marine Genomics

Cells to Shells: The genomics of mollusc exoskeletons

Mining the transcriptomes of four commercially important shellfish species for single nucleotide polymorphisms within biomineralization genes

David L.J. Vendrami [a,*], Abhijeet Shah [a], Luca Telesca [b], Joseph I. Hoffman [a]

[a] Department of Animal Behaviour, University of Bielefeld, Postfach 100131, 33501 Bielefeld, Germany
[b] Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge, Cambridgeshire CB2 3EQ, UK

ARTICLE INFO

Article history:
Received 8 October 2015
Received in revised form 17 December 2015
Accepted 23 December 2015
Available online 21 January 2016

Keywords:
Crassostrea gigas
Mya truncata
Mytilus edulis
Pecten maximus
Non-model organism

ABSTRACT

Transcriptional profiling not only provides insights into patterns of gene expression, but also generates sequences that can be mined for molecular markers, which in turn can be used for population genetic studies. As part of a large-scale effort to better understand how commercially important European shellfish species may respond to ocean acidification, we therefore mined the transcriptomes of four species (the Pacific oyster *Crassostrea gigas*, the blue mussel *Mytilus edulis*, the great scallop *Pecten maximus* and the blunt gaper *Mya truncata*) for single nucleotide polymorphisms (SNPs). Illumina data for *C. gigas*, *M. edulis* and *P. maximus* and 454 data for *M. truncata* were interrogated using GATK and SWAP454 respectively to identify between 8267 and 47,159 high quality SNPs per species (total = 121,053 SNPs residing within 34,716 different contigs). We then annotated the transcripts containing SNPs to reveal homology to diverse genes. Finally, as oceanic pH affects the ability of organisms to incorporate calcium carbonate, we honed in on genes implicated in the biomineralization process to identify a total of 1899 SNPs in 157 genes. These provide good candidates for biomarkers with which to study patterns of selection in natural or experimental populations.

1. Introduction

Climate change is one of the major factors threatening biodiversity in the 21st century (Millennium Ecosystem Assessment, 2005) and may even surpass habitat destruction as the greatest stressor in the coming decades (Leadley, 2010). Climate change is affecting the distributions, abundances, behaviors, physiologies and phenologies of many organisms (Bradshaw and Holzapfel, 2006, Forcada and Hoffman, 2014, Franks and Hoffmann, 2012). The sheer pace of ongoing change has led to mounting concerns over whether species will be able to adapt fast enough to survive (Hoffmann and Sgrò, 2011, Parmesan, 2006, Shaw and Etterson, 2012).

One important and pervasive consequence of climate change is ocean acidification (Howes et al., 2015). Atmospheric CO_2 levels have risen from 280 to >390 ppm since the onset of the industrial revolution (http://www.esrl.noaa.gov/gmd/ccgg/trends/ last accessed: October 7 2015) and average surface ocean pH has fallen from 8.16 to 8.05 over the same period (Cao and Caldeira, 2008), leading to a concomitant reduction in the availability of carbonate ions. This affects the ability of organisms such as molluscs to incorporate calcium carbonate, which is essential for building and maintaining robust skeletons and shells (Doney et al., 2009).

Genetic and genomic studies are instrumental to understanding the impacts of ocean acidification and the mechanisms by which various species might adapt to changing oceanic pH (Pespeni et al., 2013, 2012). Approaches that exploit recent advances in high-throughput sequencing, such as transcriptional profiling, are particularly powerful as they can generate vast amounts of sequence data without the need for prior genomic resources (Ekblom and Galindo, 2011). The resulting data can provide insights into patterns of gene expression among other things, and can also be used to develop single nucleotide polymorphisms (SNPs) for use in population genetic studies. Custom SNPs can be genotyped in their tens to millions, depending on the specific assay used, but most technologies rely on allele-specific oligonucleotide probes designed from the SNP flanking sequences (Qian et al., 2015).

Crassostrea gigas (the Pacific oyster), *Mytilus edulis* (the blue mussel), *Pecten maximus* (the great scallop) and *Mya truncata* (the blunt-gaper clam), are four non-model bivalve species whose shells show differences in composition with respect to the amounts of the two calcium carbonate polymorphs (calcite and aragonite), microstructure crystal size and organic matrix content (Checa and Rodríguez-Navarro, 2005; Griesshaber et al., 2013, Taylor et al., 1969). Despite their ecological and economic importance within the European Union and wider afield, relatively little is known about the capacity of these species to respond

* Corresponding author at: Department of Animal Behavior, University of Bielefeld, Postfach 100131, 33501 Bielefeld, Germany.
E-mail address: david.vendrami@student.unife.it (D.L.J. Vendrami).

http://dx.doi.org/10.1016/j.margen.2015.12.009
1874-7787/

to environmental change. An important first step has been to sequence their mantle (the tissue that secretes the shell) transcriptomes (Sleight et al., 2016–in this issue; Yarra et al., 2016–in this issue) in order to help illuminate the mechanisms underlying shell formation and maintenance through the analysis of patterns of differential gene expression. However, the same data also lend themselves to SNP discovery. Large genome-wide distributed panels of SNPs would be useful for population and quantitative genetic studies, while SNPs mined within genes specifically relating to biomineralization could be used in a candidate gene approach to explore the genetic basis of phenotypic variability and to elucidate patterns of selection in relation to prevailing environmental conditions.

Here we mined the transcriptomes of the four species using stringent criteria to identify many tens of thousands of high-quality SNPs residing within functionally annotated transcripts. Through comparisons with other studies, we then identified SNPs located within genes known to play important roles in the biomineralization process. These markers could be used for genetic studies of the four species.

2. Material and methods

2.1. Raw data

Details of the transcriptomic data that we mined for SNPs are provided in Table 1. For *C. gigas*, *M. edulis* and *P. maximus*, we used Illumina HiSeq data generated by Yarra et al. (2016–in this issue) whereas for *M. truncata* we used 454 data generated by Sleight et al. (2016–in this issue). Briefly, transcriptomes were assembled by Yarra et al. (2016–in this issue) for the first three species using between 13 and 14 unrelated, wild-caught individuals per species that were subjected to a shell repair experiment, in which experimental individuals had holes drilled in their shells and transcriptional profiling of the mantle tissue was subsequently carried out. For *M. truncata*, RNA from the mantle tissues of 9 unrelated wild-caught individuals was sequenced. The raw sequence reads and assembled contigs are available both in the original papers and via SRA (accession number: SRP064949) and http://bit.ly/1QcFiVH (*M. truncata*) and via SRA (accession number: SRP067223) and http://molluscdb.afterparty.bio.ed.ac.uk (*C. gigas*, *M. edulis* and *P. maximus*) respectively. The transcriptomes of Yarra et al. (2016–in this issue) have already been stringently quality filtered by the author, with all reads trimmed to remove low-quality base calls. We therefore conducted a quality control check on the *M. truncata* data using the program FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc last accessed: October 7 2015). Afterwards, we trimmed away the last bases of the reads with Phred scores below 30 and discarded reads shorter than 40 bp (after trimming) using FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit last accessed: October 7 2015).

2.2. SNP discovery in C. gigas, M. edulis and P. maximus

We used the Bowtie2 package (Langmead and Salzberg, 2012) to index the previous *de novo* assembled transcriptomes, align the raw reads back to the respective transcriptomes and add read groups. The resulting SAM files were then merged together and processed using Samtools (Li et al., 2009) and Picard (http://broadinstitute.github.io/picard last accessed: October 7 2015) to generate a unique indexed BAM file from which duplicate reads were removed. Subsequently, the Genome Analysis Toolkit (GATK, McKenna et al., 2010) was applied to detect variant sites with high confidence as follows: firstly, we used 'RealignerTargetCreator' and 'IndelRealigner' to identify areas containing Indels and perform a realignment step. Then, 'HaplotypeCaller' was used to create a vcf file containing a list of all of the sites recognized either as Indels or as SNPs as well as relevant information concerning the accuracy of each call. Finally, we used 'VariantsToTable' to export the data into a tab delimited file, and filtered the resulting SNPs within R (R Core Team, 2015) to retain only high-quality, informative SNPs. Specifically, we retained only SNPs that met the following criteria: total number of alleles in called genotypes (AN) ≥20, minor allele frequency (MAF) ≥0.1, Phred-scaled *p*-value using Fisher's exact test to detect strand bias (FS) ~0, depth of coverage (DP) ≥8, Z-score from Wilcoxon rank sum test of alternative *versus* reference base qualities (BaseQRankSum) ≤−1.96 or ≥1.96, root mean square of the mapping quality of reads across all samples (MQ) ≥40 and variant confidence normalized by unfiltered depth of variant samples (QD) ≥5. Finally, we evaluated the SNP parameter space for each set of called SNPs by computing a two dimensional Kernel density estimation using the 'kde2d' function in the R package MASS (Venables and Ripley, 2002). This approach uses a smoothing function to estimate a probability distribution in two dimensions (MAF and DP) for data visualization.

2.3. SNP discovery in M. truncata

We used SWAP454 (Brockman et al., 2008), which first maps the raw reads back to the assembled contigs and then determines, while taking in account an error model for the 454 data, which positions are called as SNPs according to two user-specified thresholds. The first of these, 'MIN_RATIO' corresponds to the percentage of reads that differ from the reference sequence at a given position and the second, 'MIN_READS' to the number of copies present of the minor allele. To minimize the possibility of false positives arising from sequencing error, MIN_RATIO was set to 0.1 and MIN_READS was set to 5. The SNP parameter space was evaluated as described for the other three species.

2.4. Flanking sequences

We used the BEDtools software package (Quinlan, 2014) to extract the flanking sequences of each SNP (60 bp on either side) for probe design. We discarded SNPs with <50 bp of flanking sequence due to proximity to the start or end of contigs.

2.5. Sequence annotation

For each species, we filtered the original transcriptome to create a subset FASTA file comprising only those contigs within which at least one SNP was found. We performed blastx sequence similarity searching of these contig sequences against the GenBank nr and Swiss-Prot protein databases with a standard e-value cutoff of $1e^{-10}$. When a search produced more than one result for a given contig, only the annotation with the lowest e-value was retained. Finally, we retrieved Gene Ontology (GO) annotations for each contig. These operations were performed as described by De Wit et al. (2012). Subsequently, we classified GO annotations according to three major functional categories: biological processes, molecular function and cellular component. The Web Gene Ontology Annotation Plotting tool (Ye et al., 2006) was then utilized to obtain GO slim annotations and these were used to plot the distribution (%) of gene ontology terms among the annotated unique

Table 1
Summary of the raw data available for the four shellfish species.

Species	Literature reference	Type of data	Number of individuals sequenced	Collection site (s)	Number of reads
C. gigas	Yarra et al. (2016-in this issue)	Illumina paired end (125 + 125 bp)	13	Barmore Bay (Scotland, UK)	208.1 million
M. edulis	Yarra et al. (2016-in this issue)	Illumina paired end (125 + 125 bp)	14	Tarbert (Scotland, UK)	286 million
P. maximus	Yarra et al. (2016-in this issue)	Illumina paired end (125 + 125 bp)	14	Eilean Buidhe Island (Scotland, UK)	180.4 million
M. truncata	Sleight et al. (2016-in this issue)	454 single end (500 bp)	9	Dunstaffnage Bay (Scotland, UK)	702,006

sequences. Specifically, we employed the GO slim set archive supplied by the Web Gene Ontology Plotting tool, downloaded from the GO FTP archive.

2.6. Identification of candidate SNPs

Finally, in order to identify SNPs within biomineralization-related genes, we compared the above annotations with lists of proteins implicated in the biomineralization process by previous studies (see Table 2 and Supplementary Table 1 for details). Specifically, we searched for the protein names and GenBank protein accession numbers against our annotations. Any contig annotation matching perfectly with one of these terms was flagged as being putatively involved in the biomineralization process.

3. Results

3.1. SNP discovery

We interrogated the transcriptomes of the four shellfish species using GATK for the Illumina datasets (*C. gigas*, *M. edulis* and *P. maximus*) and SWAP454 for the 454 dataset (*M. truncata* see Table 3 for details). GATK recovered between 892,753 and 2,001,761 SNPs per species, which after stringent quality filtering was reduced to 30,132–47,159 SNPs (Table 3). SWAP454 identified only 8267 high quality SNPs in *M. truncata* (Table 3) consistent with many fewer bases having been assembled for this species. The majority of SNPs in all four species (95–98%) had at least 50 bp of flanking sequence on either side (Table 3) and are therefore suitable for designing probes.

To explore the SNP-calling parameter space for each species, we computed two dimensional Kernel density estimates. The results for *C. gigas*, *M. edulis* and *P. maximus* appear highly similar to each other (Fig. 1a–c) and reveal peaks in the number of SNPs identified at a MAF of around 0.125 and a log depth of coverage of approximately 1.75, which corresponds to around 56× depth of coverage. In contrast, *M. truncata* is characterized by a peak in the number of SNPs identified at a MAF of around 0.45 and a log depth of coverage of approximately 1.1 (Fig. 1d), which corresponds to around 12.6× depth of coverage.

3.2. Sequence annotation

For each species, contigs containing at least one high quality SNP with ≥50 bp of flanking sequence were annotated relative to the nr and Swiss-Prot protein databases (see Supplementary Tables 2–5 for details of the SNPs together with relevant annotations). Table 4 shows the proportion of contigs containing SNPs that were annotated for each of the four species. To provide an overview, we used GO slim annotations to classify each contig according to three major functional categories: biological process, molecular function and cellular component. These results are summarized in Fig. 2, which not only reveals a broad range of classifications, but also shows a high level of consistency in the relative frequencies of the various classifications across the four species.

3.3. Identification of candidate SNPs

Finally, we sought to identify SNPs residing within genes implicated in biomineralization by cross-referencing our annotations to proteins listed in other studies. A total of 494 contigs containing 1899 SNPs revealed exact matches to 157 out of a total of 643 proteins. A breakdown by species is provided in Table 2 and Supplementary Table 1, while the individual SNPs are given in Supplementary Tables 2–5. Fig. 3 shows that there was considerable overlap among the four species in the specific biomineralization related proteins that contained SNPs.

4. Discussion

We mined the transcriptomes of four European shellfish species to identify SNPs residing within functionally annotated transcripts. Even after applying fairly stringent criteria, we identified over 121,000 SNPs in total. Although not strictly comparable, the larger Illumina datasets available for *C. gigas*, *M. edulis* and *P. maximus* yielded many more SNPs than the *M. truncata* 454 dataset. The far greater depth of coverage provided by Illumina sequencing also facilitated the discovery of lower frequency SNPs. Regardless of the sequencing technology, we were also able to identify panels of SNPs within genes implicated in the biomineralization process in all four species. These putative markers provide a resource for future candidate gene studies.

This study exploited two types of sequence data, Illumina HiSeq and 454. A direct comparison of the two methods is not possible because none of the species were sequenced with both technologies. We also used different SNP calling programs for the two types of data, partly because GATK cannot cope with 454 data, but also because the respective SNP callers need to take into account specific error models for the data. Nevertheless, it is fairly obvious that the Illumina datasets are superior in respect of SNP discovery. First, the larger Illumina datasets yielded between 3.6 and 5.7 times more high-quality SNPs than the 454 dataset, even after applying very stringent quality criteria. Second, two dimensional Kernel density estimates reveal marked differences in the SNPs called from the Illumina and 454 datasets, the former being characterized mainly by low frequency SNPs sequenced at high coverage, and the latter by high frequency SNPs sequenced at lower coverage. This could potentially have downstream consequences because the inclusion of highly polymorphic SNPs in a genotyping assay can affect allele frequency estimates, upwardly bias estimates of genetic diversity and generate false bottleneck signatures (Morin et al., 2004). However, this can be difficult to judge as *in silico* estimates of MAF are often based on a small discovery pool of individuals and therefore tend to correlate poorly if at all with the true MAF in a larger population (Hoffman et al., 2012).

When selecting a panel of SNPs for inclusion in a genotyping project it is also important to bear in mind the diversity of currently available genotyping technologies. Small to medium throughput technologies include Applied Biosystem's SNPlexTM and TaqMan® SNP genotyping assays, Sequenom's iPlex® assay, Beckman Coulter's SNPstream® and LGC's KASP™ assay. At the opposite end of the spectrum are Illumina Infinium iSelect® and Affymetrix Axiom® arrays, which can support several thousands to millions of SNPs. All of the above technologies rely on probes designed from SNP flanking sequences, although the

Table 2
Numbers of proteins implicated in the biomineralization process revealing matches to the contig annotations. The last row shows the total numbers of SNPs with the numbers of contigs (both shown in parentheses) and proteins within which they reside. Proteins that are listed in more than one of the five studies only count towards the total once.

Species	Reference	No. of genes	*C. gigas*	*M. edulis*	*P. maximus*	*M. truncata*
Pinctada maxima and *Haliotis asinina*	Jackson et al. (2009)	57	13	10	13	6
Laternula elliptica	Clark et al. (2010)	347	37	29	17	5
Pinctada margaritifera	Joubert et al. (2010)	80	6	7	6	6
Patinopecten yessoensis	Sun et al. (2015)	167	51	50	39	15
M. truncata	Sleight et al. (in this issue)	7	2	2	2	2
Total	–	643	101 (714; 171)	92 (564; 153)	69 (383; 118)	28 (238; 52)

Table 3
Numbers of SNPs mined from the transcriptomes of the four species (see Section 2 for details).

Species	No. of contigs (bases assembled)	SNP caller	Total no. of SNPs	No. of high quality SNPs	No. of high quality SNPs with > 50 bp flanking sequence	No. of contigs containing at least one SNP with >50 bp flanking sequence	Average no. of high quality SNPs per kb
C. gigas	426,020 (253,581,654)	GATK	1,474,197	47,159	46,555	12,024	0.07
M. edulis	559,818 (282,990,328)	GATK	2,001,761	35,495	34,943	10,175	0.04
P. maximus	228,088 (154,009,095)	GATK	892,753	30,132	29,536	10,662	0.10
M. truncata	20,106 (13,576,161)	SWAP454	–	8267	7824	1855	0.42
Total	1,234,032 (704,157,238)	–	–	121,053	118,858	34,716	–

exact number of base pairs required varies with the method. To provide a rough estimate of the number of SNPs with enough flanking sequence to design probes, we therefore filtered our dataset of high quality SNPs for those with at least 50 bp of sequence on either side. The proportion retained was very high (between 95 and 98%) indicating that most of the SNPs should be suitable for probe design. However, it is usual to conduct an additional filtering step in which proprietary algorithms are used to analyze the probe sequences for compatibility with a specific genotyping technology. This typically leads to a fraction of SNPs being flagged as unlikely to genotype successfully.

To obtain an impression of the functional diversity of the genes from which SNPs were mined, we annotated the relevant contigs with respect to the GenBank nr and Swissprot protein databases and then used GO slim to generate a summary. A diverse range of biological processes, molecular functions and cellular components were represented in all four of the species. This is to be expected given the large size of the transcriptomic resources involved. In addition, the frequency distributions of the various classifications are very similar across the four species. This probably reflects species similarities, although it is also possible that deriving the annotations from the same databases could have led to an increase in the apparent level of similarity.

To elucidate patterns of selection in relation to ocean acidification, it would be desirable to identify SNPs residing within genes involved in biomineralization. We therefore cross-referenced our annotations to other studies to identify over 1800 SNPs residing within almost 500 contigs corresponding to 157 candidate genes. These include perlucin, a protein that regulates pearl formation and shell biomineralization by promoting calcium carbonate precipitation and calcite crystal modification (Wang et al., 2008), chitin synthase, the enzyme responsible for chitin deposition in molluscan shells and nacre (Weiss et al., 2006), carbonic anhydrase, which is known to be implicated in matrix mineralization by generating an acidic environment (Clark et al., 2010) and many others.

An important caveat to this paper, as well as most other SNP discovery efforts, is that mining SNPs from a small number of individuals (in this case between 9 and 14 individual per species) is unlikely to capture all of the genetic variation within a population. However, all of the samples used for this study comprised ostensibly unrelated wild-caught animals and did not originate from cultured lines. Therefore, the published

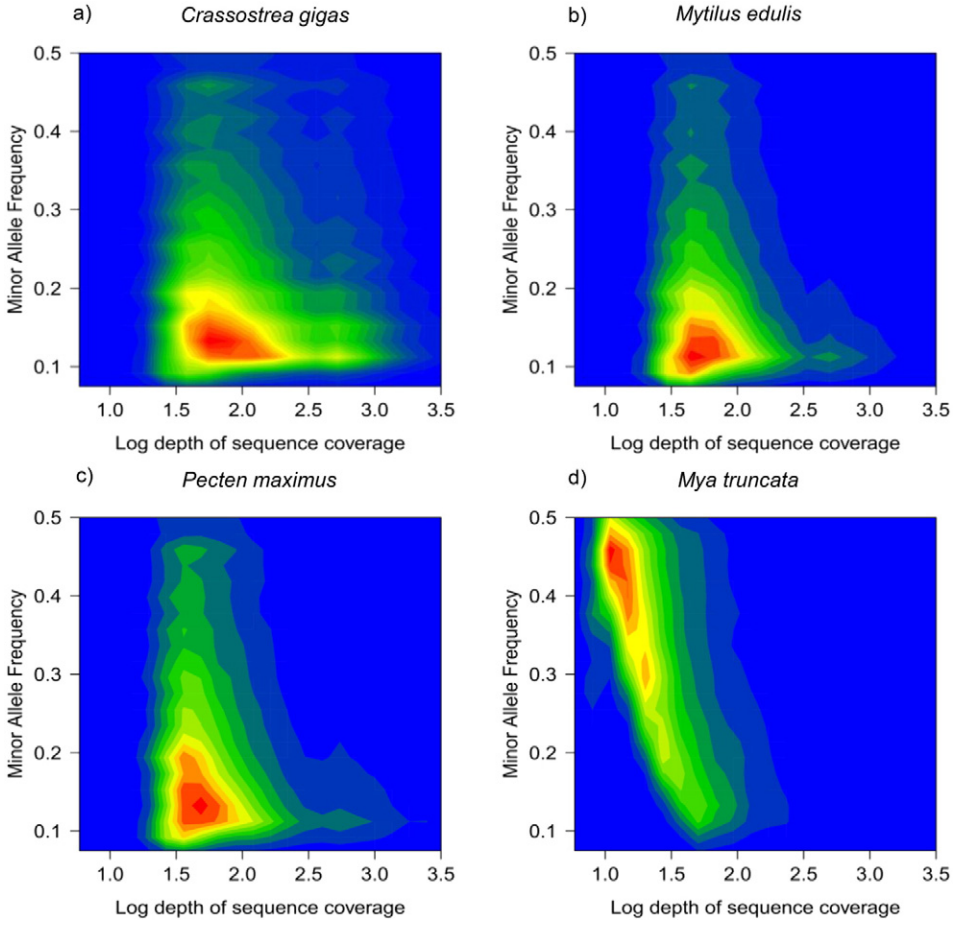

Fig. 1. Variation in the number of SNPs discovered in the four species with MAF and the logarithm of the depth of sequence coverage. The area shaded in red indicates the parameter space associated with the greatest number of SNPs discovered.

Table 4
The numbers and proportions of SNP containing contigs for which nr, Swiss-Prot and Gene Ontology annotations could be retrieved.

Species	nr/Swissprot	Gene ontology
C. gigas	9160 (76.17%)	5704 (47.43%)
M. edulis	6528 (64.16%)	4908 (48.24%)
P. maximus	4460 (41.86%)	3347 (31.4%)
M. truncata	753 (40.57%)	643 (34.64%)

transcriptomes represent natural diversity as far as is possible given financial and other constraints. Our study identified variants with a range of allele frequencies going down to 0.1. However, it should be born in mind that alleles present at lower frequencies will have been missed. This is a common problem with SNP studies and the use of SNP panels derived from a small discovery pool of individuals can potentially lead to inflated estimates of genetic diversity (Lynch et al., 2014). One way around this is to use approaches like restriction site associated DNA (RAD) sequencing, which can potentially capture a more representative array of variants (Davey et al., 2011). However, like most methods, RAD sequencing has its own issues and potential biases (Davey et al., 2013).

The SNPs we have identified provide a resource for future population genetic studies of the four shellfish species, while also opening up avenues for exploring how selection may be acting at genes involved in

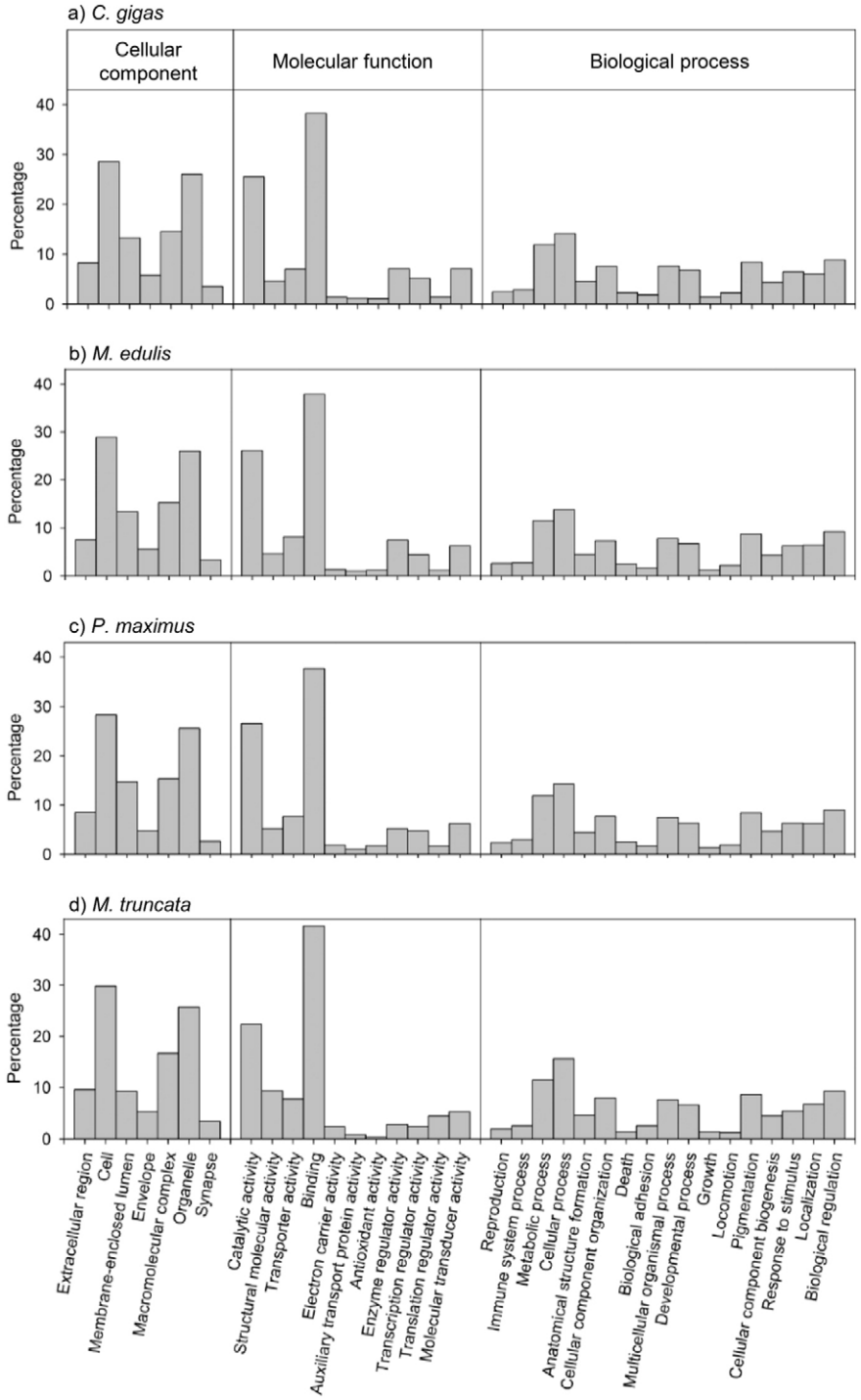

Fig. 2. Gene ontology classifications of SNP-containing contigs for the four shellfish species summarized for three major categories: biological process, molecular function and cellular component.

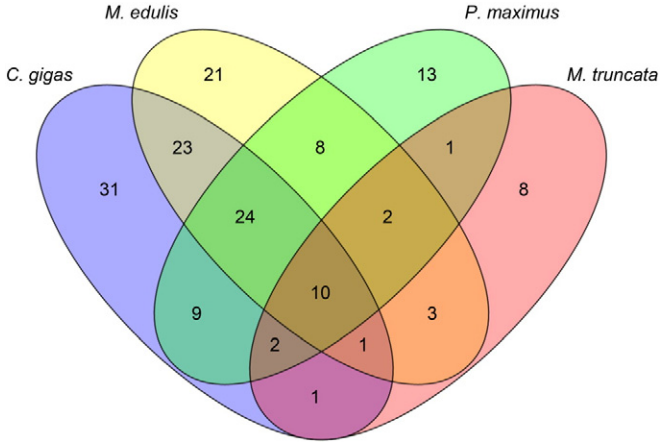

Fig. 3. Venn diagram showing overlaps among the four shellfish species in biomineralization related proteins containing at least one SNP.

biomineralization. One popular approach for distinguishing neutral loci from those subject to selection exploits the fact the latter often show different, 'outlier' patterns of variation (Beaumont and Balding, 2004). Within this framework it would be possible to design a study to test for such patterns, for example, across environmental clines. To do this it would be necessary to screen randomly selected neutral SNPs, potentially obtained by RAD sequencing, in combination with the candidate biomineralization SNPs, the prediction being that the latter should be more likely to exhibit outlier patterns. Alternatively, these SNPs could potentially be used as 'biomarkers' with which to document allele frequency changes at key genes, either in wild populations or in relation to an experimental challenge. Again, it could be helpful in many cases to combine both neutral and candidate SNPs in order to provide insights into both neutral and adaptive processes.

The transcriptome and raw reads (*M. truncata*) are available at http://bit.ly/1QcFiVH and via SRA (accession number: SRP064949) respectively.

The transcriptome and raw reads (*C. gigas*, *M. edulis*, *P. maximus*) are available at http://molluscdb.afterparty.bio.ed.ac.uk and via SRA (accession number: SRP067223) respectively.

The SNPs are available at the dbSNP (release June 2016, http://www.ncbi.nlm.nih.gov/SNP/) submitted SNP (ss) numbers: 1966566540 - 1966613094 (*C. gigas*), 1966531597 - 1966566539 (*M. edulis*), 1966613328 - 1966644512 (*P. maximus*) and 1966613095 - 1966650454 (*M. truncata*).

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.margen.2015.12.009.

Acknowledgments

We thank Victoria A. Sleight and Michael A.S. Thorne for providing transcriptome sequence data for *M. truncata* and Tejaswi Yarra for providing transcriptomes and cleaned raw reads for the other three species. We also thank Melody S. Clark and Elizabeth M. Harper for proof reading the manuscript. The research leading to these results has received funding from the European Union Marie Curie Seventh Framework Programme under grant agreement no. 605051.

References

Andrews, S., 2015. http://www.bioinformatics.babraham.ac.uk/projects/fastqc (last accessed: October 7) .
Beaumont, M.A., Balding, D.J., 2004. Identifying adaptive genetic divergence among populations from genome scans. Mol. Ecol. 13, 969–980.
Bradshaw, W.E., Holzapfel, C.M., 2006. Evolutionary response to rapid climate change. Science 312, 1477–1478.
Brockman, W., Alvarez, P., Young, S., Garber, M., Giannoukos, G., Lee, W.L., Russ, C., Lander, E.S., Nusbaum, C., Jaffe, D.B., 2008. Quality scores and SNP detection in sequencing-by-synthesis systems. Genome Res. 18, 763–770.
Cao, L., Caldeira, K., 2008. Atmospheric CO_2 stabilization and ocean acidification. Geophys. Res. Lett. 35.
Checa, A.G., Rodríguez-Navarro, A.B., 2005. Self-organisation of nacre in the shells of Pterioida (Bivalvia: Mollusca). Biomaterials 26, 1071–1079.
Clark, M.S., Thorne, M.A., Vieira, F.A., Cardoso, J.C., Power, D.M., Peck, L.S., 2010. Insights into shell deposition in the Antarctic bivalve *Laternula elliptica*: gene discovery in the mantle transcriptome using 454 pyrosequencing. BMC Genomics 11, 362.
Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12, 499–510.
Davey, J.W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., Blaxter, M.L., 2013. Special features of RAD sequencing data: implications for genotyping. Mol. Ecol. 22, 3151–3164.
De Wit, P., Pespeni, M.H., Ladner, J.T., Barshis, D.J., Seneca, F., Jaris, H., Overgaard, T.N., Morikawa, M., Palumbi, S.R., 2012. The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. Mol. Ecol. Resour. 12, 1058–1067.
Doney, S.C., Fabry, V.J., Feely, R.A., Kleypas, J.A., 2009. Ocean acidification: the other CO_2 problem. Mar. Sci. 1.
Ekblom, R., Galindo, J., 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107, 1–15.
Forcada, J., Hoffman, J.I., 2014. Climate change selects for heterozygosity in a declining fur seal population. Nature 511, 462–465.
Franks, S.J., Hoffmann, A.A., 2012. Genetics of climate change adaptation. Genetics 46.
Griesshaber, E., Schmahl, W.W., Ubhi, H.S., Huber, J., Nindiyasari, F., Maier, B., Ziegler, A., 2013. Homoepitaxial meso-and microscale crystal co-orientation and organic matrix network structure in *Mytilus edulis* nacre and calcite. Acta Biomater. 9, 9492–9502.
Hannon Lab, 2015. http://hannonlab.cshl.edu/fastx_toolkit (last accessed: October 7) .
Hoffman, J.I., Tucker, R., Clark, M.S., Forcada, J., Slate, J., 2012. Rates of assay success and genotyping error when single nucleotide polymorphism genotyping in non-model organisms: a case study in the Antarctic fur seal. Mol. Ecol. Resour. 12, 861–872.
Hoffmann, A.A., Sgrò, C.M., 2011. Climate change and evolutionary adaptation. Nature 470, 479–485.
Howes, E.L., Joos, F., Eakin, M., Gattuso, J., 2015. An updated synthesis of the observed and projected impacts of climate change on the chemical, physical and biological processes in the oceans. Front. Mar. Sci. 2, 36.
Jackson, D.J., McDougall, C., Woodcroft, B., Moase, P., Rose, R.A., Kube, M., Reinhardt, R., Rokhsar, D.S., Montagnini, C., Joubert, C., Piquemal, D., Degnan, B.M., 2009. Parallel evolution of nacre building gene sets in molluscs. Mol. Biol. Evol. (msp278).
Joubert, C., Piquemal, D., Marie, B., Manchon, L., Pierrat, F., Zanella-Cléon, I., Cochennec-Laureau, N., Gueguen, Y., Montagnani, C., 2010. Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: focus on biomineralization. BMC Genomics 11, 613.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.
Leadley, P., 2010. Biodiversity Scenarios: Projections of 21st Century Change in Biodiversity, and Associated Ecosystem Services: a Technical Report for the Global Biodiversity. Outlook 3 (No. 50). UNEP/Earthprint.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.
Lynch, M., Bost, D., Wilson, S., Maruki, T., Harrison, S., 2014. Population-genetic inference from pooled-sequencing data. Genome Biol. Evol. 6, 1210–1218.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.
Millennium Ecosystem Assessment, 2005. Ecosystems and Human Well-being: Biodiversity Synthesis. World Resources Institute, Washington, DC.
Morin, P.A., Luikart, G., Wayne, R., 2004. SNP's in ecology, evolution and conservation. Trends Ecol. Evol. 19, 208–216.
Parmesan, C., 2006. Ecological and evolutionary responses to recent climate change. Annu. Rev. Ecol. Evol. Syst. 637-669.
Pespeni, M.H., Garfield, D.A., Manier, M.K., Palumbi, S.R., 2012. Genome-wide polymorphisms show unexpected targets of natural selection. Proc. R. Soc. Lond. B Biol. Sci. 279, 1412–1420.
Pespeni, M.H., Barney, B.T., Palumbi, S.R., 2013. Differences in the regulation of growth and biomineralization genes revealed through long-term common-garden acclimation and experimental genomics in the purple sea urchin. Evolution 67, 1901–1914.
Qian, X., Pu, D., Liu, B., Xiao, P., 2015. Effect of oligonucleotide probes substituted by deoxyinosines on the specificity of SNP detection on the DNA microarray. Electrophoresis 36, 263–270.
Quinlan, A.R., 2014. BEDTools: the Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinform. 11-12.
R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Shaw, R.G., Etterson, J.R., 2012. Rapid climate change and the rate of adaptation: insight from experimental quantitative genetics. New Phytol. 195, 752–765.
Sleight, V.A., Thorne, M.A.S., Peck, L.S., Arivalagan, J., Berland, S., Marie, A., Clark, M.S., 2016. Characterisation of the mantle transcriptome and biomineralisation genes in the blunt-gaper clam, *Mya truncata*. Mar. Genomics 27, 47–55.
Sun, X., Yang, A., Wu, B., Zhou, L., Liu, Z., 2015. Characterization of the mantle transcriptome of Yesso scallop (*Patinopecten yessoensis*): identification of genes potentially involved in biomineralization and pigmentation. PLoS One 10.
Tans, P., Keeling, R., 2015. www.esrl.noaa.gov/gmd/ccgg/trends/ (last accessed: October 7) .

Taylor, J.D., Kennedy, W.J., Hall, A., 1969. The shell structure and mineralogy of the Bivalvia Introduction. Nuculacea–Trigonacea. Bull. Br. Mus. (Nat. Hist.) Zool. 3 (Supp. l), 1–125.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S. fourth ed. Springer, New York.
Wang, N., Lee, Y.H., Lee, J., 2008. Recombinant perlucin nucleates the growth of calcium carbonate crystals: molecular cloning and characterization of perlucin from disk abalone, *Haliotis discus discus*. Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 149, 354–361.
Weiss, I.M., Schönitzer, V., Eichner, N., Sumper, M., 2006. The chitin synthase involved in marine bivalve mollusk shell formation contains a myosin domain. FEBS Lett. 580, 1846–1852.
Wysoker, A., Tibbetts, K.N., Fennell, T., 2015. http://broadinstitute.github.io/picard (last accessed: October 7) .
Yarra, T., Gharbi, K., Blaxter, M., Peck, L., Clark, M.S., 2016. Characterization of the mantle transcriptome in bivalves: *Pecten maximus*, *Mytilus edulis*, and *Crassostrea gigas*. Mar. Genomics 27, 9–15.
Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., Wang, J., 2006. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 34 (Suppl. 2), W293–W297.

David Vendrami is a PhD student with interests in population genetics and genomics. He earned both his Bachelor's and Master's Degree at the University of Ferrara, Italy. He spent one year at the NOAA's Southwest Fisheries Science Center in Santa Cruz (California, USA) to write the thesis for his Master's Degree. Currently, he is enrolled as a PhD student at the University of Bielefeld as part of the CACHE network.

Abhijeet Shah is a researcher whose activity is particularly focused on genomics and bioinformatics. After earning his Bachelor's Degree at the Iowa State University of Science and Technology in Ames (Iowa, USA) and his Master's Degree at the Munich Graduate School for Evolution, Ecology and Systematics in Munich (Germany), he worked as a short-time scientist in Manchester, Paris and Helsinki. Currently he is a research assistant at the University of Bielefeld in Germany.

Luca Telesca is a marine biologist with a broad interest in the effect of environmental changes on marine benthic communities. After completing a BSc in Biology and a MSc in Marine Sciences at the University of Rome "La Sapienza", he has started a PhD in the Department of Earth Sciences at the University of Cambridge working on the natural variability of commercial shellfish (bivalves) and the effect of environmental changes on their shells structure.

Joe Hoffman is a population geneticist with broad interests in the population structure and genetic diversity of marine organisms. After completing his PhD at the Department of Zoology at Cambridge university, he held a series of fellowships before moving to Bielefeld University in Germany, where he currently heads the behavioral genetics laboratory.