Recent advances in EST sequencing in Crassostrea gigas: towards the sequencing of the Pacific oyster genome.

P Boudry¹, A Huvet², S Lapègue¹, C Sauvage¹, A Tanguy³, D Moraga⁴, A Goostrey⁵, P Prunet⁶, V Boulo⁷, Y Gueguen⁷, J Moal², E Fleury², JF Samain², M Mathieu², P Favrel²

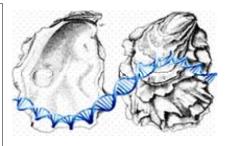
¹LGP Ifremer La Tremblade, ²UMR PEMM Ifremer Brest/Univ. Caen, ³CNRS-Université Paris VI, Roscoff, ⁴LEMAR, University of Brest, ⁵Marine Laboratory, Plymouth, 6SCRIBE, INRA, Rennes, 7UMR 5119 Ifremer-CNRS-Université Montpellier II.

Introduction

Rationale for EST and genome sequencing of the Pacific oyster:

- its membership of the Lophotrochozoa, an little studied branch of the Eukaryotes,
- its worldwide aquacultural importance (4.2 million metric tons, worth 3.5 billion US dollars),
- its high fecundity, with concomitantly high DNA polymorphism,
- its use as a biosensor of coastal pollution.

In 2004, an international community of biologists, the Oyster Genome Consortium (OGC) led by D. Hedgecock (University of Southern California) and uniting 70 participants from 10 countries, first presented the Pacific oyster Crassostrea gigas as a genome-sequencing candidate to Joint Genome Institute (JGI, USA) (Hedgecock et al., 2005). Besides this project (that remains to be accepted), several EST sequencing projects have been successfully initiated at JGI, Genoscope (France) and the Max Planck Institute (Germany). These will considerably enlarge the first public database specifically dedicated to C. gigas (Gueguen et al., 2003; www.ifremer.fr/GigasBase).



Genome size and characteristics

Consensus haploid genome size is ~0.89 pg or 824 Mb. Genome sizes of 174 mollusks range from 0.43pg (for the limpet Lottia gigantea) to 5.88pg (for the whelk Neobuccinum eatoni), with a mean of 1.80 ± 0.07 pg. Among those, only 23 mollusks have genome sizes smaller than that of the Pacific oyster.

Protein polymorphism of marine bivalves is among the highest for animals. In the Pacific oyster, the average density of SNPs was estimated to be among the highest levels reported to date with one SNP every 61 bp in coding regions and one every 41 bp in non-coding regions (Sauvage et al., submitted). As a result, abundant non-amplifying PCR-null alleles are commonly observed. Because of the enormous fecundity and the very large population sizes in this species, a high genetic load is maintained in wild populations, explaining (1) frequent distorted inheritance ratios in lab-reared progeny of wild parents and (2) correlation of heterozygosity with fitness related traits in natural populations.

The available microsatellite and AFLP-based linkage maps (Hubert and Hedgecock 2004 ; Yu and Guo 2003), have 10 linkage groups, in accord with haploid chromosome number and cover ~80% of the Pacific oyster genome. SNPs are now being mapped and a physical map is being developed using a BAC (Cunningham et al., 2006) fingerprinting approach.

Current EST sequencing projects and their use in gene expression studies

More than 8000 ESTs resulting from 12 subtractive and 3 normalized libraries from 6 tissues have been sequenced by MPI within the framework of the EU funded projects Aquafirst and Marine Genomics Europe. In addition, normalized tissue or stage specific libraries are currently being sequenced by Genoscope and JGI.

A new EST database is hosted by INRA on the Sigenae Information system, bringing together these new ESTs with all previously published ESTs and mRNAs. The 1758 obtained contigs plus 7518 singletons are now being used to construct an European C. gigas microarray slide.

and half of the all and half of the second s					Current Sigense Contigs Dyster Version							approbilation for the town to 1 - 480	E Cantig sypmitti	8602105r1_m12rev.1.a.cg.1			
					Electoring Persion	Number of Sequences	Coolings CAP3	tumber of Singlets	Contign Obtained Du	Libraries fo	Blast	···· Drashtenikana •• Trashtenikana		amontari			
	Ausselleret Banken Brok B	unt bea	613-000	Treed .		Raw assembly		1758		2000-047			· Association allow	E Detailed view			
-				11 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1									··· Mid-reality		Paskins + DAI	Closent + Report + Decembers + Signi + Improve	
	Review engements representing		Ander 201	1994		and setterild							 Provide Children Provide Children 		owner to implice the	namiPhiliphia at heat as \$	- Hateat
	Therease Review and an analysis		August (1988	100		Cont.actuality							··· Loynet neuronau to 10020		-		The second se
()	Berter teas Lategory		1900 2014		100.00							AR	Line Ermanisi ta		- Marine	T STATE	
er ited	meners 4				INRA							TEULOUSE	· Search Code Streeter			I	10 20 10 ID
	AqueFirst Date Access				dh.n.a.r					4-944		PEROPOTE-	G they same (Soldard)		LANSTO DITA	to favori shart di la	
	Table Description Address of Addr	1. State 1.	10 C	And a second sec									· Daniel Bare		CO-REDAR		init(albeit
				A											Office stand		and the second
		5 C.		1 Same									Dare and describeda		127488		
	Processed balthes												B Aland Carling Browner		Contraction of the second		
													· Coolig Revenue date		(Chass-Web		
		ang incoded seg is	and then any	And at a said									Other Inks		- Descartage	against the second s	and the second se
	and a construction of the second seco	-		-									- Dana		Q Base But	Contraction of the second states of the second	
	manual points (1981) 11 april 10	E (S.	1.00	-									- Appelled		CONTRACTOR AND	-	COLUMN REPORTED
	manufactures (200.071) capitals and			100									··· Station		Longes.		Stand Mead
	maintaine 200211 applet W	38 10	16.8	100									CONTRACTOR AND				
	market soles 200.07-17 coadant int	At 10.	204	100									Stanger IIIn			Sails long buser transities gips wrole, ogt, in	the input for the call of the second se

Conclusions and Future work

Past and present work on EST sequencing in Europe and the USA provide a solid basis for gene expression studies, based on SSH (eg Huvet et al., 2004) and sequencing of normalized libraries (Tanguy et al., in prep), microarrays (Jenny et al., in press), SAGE, MPSS (Hedgecock et al., 2007) and other approaches (Saavedra and Bachère, 2006).

➡ OCG is currently seeking new opportunities to have the Pacific oyster genome sequenced. A conference proposal will soon put forward to the National Research Initiative Animal Genomes (USDA).

References

- <u>References:</u>
 Cunningham et al. (2006). New resources for marine genomics: Bacterial artificial chromosome libraries for the eastern and pacific oysters (*Crassostrea virginica* and *C. gigas*) Marine Biotechnology 8: 521-533.
 Gueguen et al. (2003). Immune gene discovery by expressed sequence tags generated from hemocytes of the bacteria-challenged oyster, *Crassostrea gigas*. Gene 303: 139-145.
 Hedgecock et al. (2005). The case for sequencing the Pacific oyster genome. Journal of Shellfish Research 24: 429-441.
 Hedgecock et al. (2007). Transcriptomic analysis of growth heterosis in larval Pacific oysters (*Crassostrea gigas*). BAS 104(7): 2313-2318.
 Hubert and Hedgecock (2003). Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. Genetics 168: 351-362.
 Huvet et al. (2004). The identification of genes from the oyster *Crassostrea gigas* that are differentially expressed in progeny exhibiting opposed susceptibility to summer mortality. Gene 343(1): 211-220.
 Jenny et al. (2007). A cDNA Microarray for *Crassostrea gigas*. Marine Biotechnology (in press)
 Saavedra and Bachere (2006). Bivalve genomics. Aquaculture 256: 1-14.
 Sauvage et al. Development of SNPs and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. Submitted to Gene.
 'Tanguy et al. A contribution of "Marine Eurone" to the improvement of FST data in marine bivalves. In pren

- •Tanguy et al. A contribution of "Marine Genomic Europe" to the improvement of EST data in marine bivalves. In prep. Yu and Guo (2003). Genetic linkage map of the eastern oyster *Crassostrea virginica* Gmelin. Biological. Bulletin 204, 327-338.