



Sauvage Christopher
Laboratoire de Génétique
et Pathologie
Avenue Mus de Loup
17390, La Tremblade

DEVELOPPEMENT DE MARQUEURS MOLECULAIRES LIES A LA
RESISTANCE A LA MORTALITE ESTIVALE CHEZ L'HUITRE CREUSE
CRASSOSTREA GIGAS -APPROCHE QTL

Université de La Rochelle
U.F.R des Sciences et Technologies
Ecole doctorale ED0374

Thèse soutenue le 9 juin 2008
Spécialité Biologie

Membres du Jury :

Mr Boudry Pierre	Cadre de recherche, Ifremer	Directeur de thèse
Mme Le Roy Pascale	Directeur de recherche, INRA	Rapporteur
Mr Filip Volckaert	Professeur, Katholieke Universiteit Leuven	Rapporteur
Mme Garcia-Meunier Pascale	Professeur, Université de La Rochelle	Examineur
Mme Quillet Edwige	Directeur de recherche, INRA	Examineur
Mme Lapègue Sylvie	Cadre de recherche, Ifremer	Responsable scientifique

Remerciements

Le moment tant attendu est enfin arrivé... ! Terminer sa thèse. Cependant, je dirai que je termine plus qu'une thèse, j'arrive au terme d'une époque, d'une aventure, qui a duré plus de 5 ans (record à battre !). En effet, je suis arrivé au début de l'année 2003 à la station Ifremer de La Tremblade, comme stagiaire de maîtrise, pour travailler sur le polymorphisme des gènes de l'amylase chez l'huître creuse. Après quelques mois de cours à l'USTL, je suis revenu en 2004, pour effectuer mon stage de DEA. Je n'ai plus quitté la région et le laboratoire jusqu'à aujourd'hui, car Pierre Boudry m'a offert la possibilité de terminer mon cursus universitaire, par un doctorat.

Au cours de ces cinq années, j'ai rencontré beaucoup de personnes qui ont contribué, de près comme de loin, au bon déroulement de l'ensemble de mes travaux. C'est pourquoi, je souhaite vivement les remercier. Pour ouvrir le bal, je remercie très sincèrement tous les membres de mon jury de thèse, qui m'ont fait l'honneur de juger mon travail. Je remercie donc Pascale Le Roy de l'INRA de Rennes, Pascale Garcia de l'université de La Rochelle, Edwige Quillet de l'INRA de Jouy en Josas et enfin Filip Volckaert de la Katholieke Universiteit Leuven, en Belgique (Leve België!).

Ensuite, je remercie Philippe Gouletquer et Jean Prou qui se sont succédés au poste de chef de station et m'ont accueilli au sein du laboratoire. Mes remerciements vont également à Tristan Renault, directeur de département et chef de laboratoire, qui dirige le laboratoire avec efficacité, rigueur et humanité.

Bien évidemment, je souhaite vivement remercier ma responsable scientifique et mon directeur de thèse, Sylvie Lapègue et Pierre Boudry. Ils m'ont fait pleinement confiance et m'ont laissé une grande autonomie au cours des trois années de doctorat. Ils ont toujours été disponibles pour discuter ou m'aider dans mes démarches de toute nature. Je les remercie également pour les corrections et commentaires constructifs qu'ils ont apportés à ce manuscrit afin de l'améliorer.

Le microcosme de la station Ifremer de La Tremblade est peuplé de figures emblématiques qui le rendent si particulier et attachant. Je vais essayer de les remercier et de n'oublier personne. Si tel est le cas, j'en suis vraiment désolé...

Tout d'abord, je remercie énormément Serge, mon équipier de co-voiturage, pour l'ensemble de son œuvre ! Celle-ci reflète le travail considérable qu'il a accompli pour le génotypage des microsatellites, dans le cadre du projet Aquafirst. Au-delà de ce super boulot, je me souviendrai de Serge pour ses jeux de mots et autres blagues mais aussi pour sa bonne humeur quotidienne !

Ensuite, si je devais désigner un chef d'orchestre, je pense que Nicole peut occuper ce poste. Je n'ai pas compté le nombre de fois où je suis allé la voir pour lui demander : « Nicole, tu sais pas où je peux trouver ... », mais ce chiffre doit être énorme ! Je te remercie Nicole pour le travail de fourmi que tu as accompli dans la gestion et la réorganisation du labo, pour l'aide que tu m'as apporté mais aussi pour

nos discussions, accoudés à la paillasse. Trop souvent, nous, les thésards, te faisons tourner en rond avec toutes les conneries qui nous passent par la tête !

Depuis, plus de trois ans, je partage le bureau avec Abdellah. Je le remercie pour ses conseils, notamment scientifiques, et nos discussions ! Il m'a initié à la cytogénétique moléculaire, au « Free Jazz », à la cuisine marocaine, mais surtout au meilleur vin de France, le vin de Bourgogne. « Je te remercie Mec et te souhaite bon courage pour la suite ! »

J'ai également côtoyé beaucoup d'autres personnes avec qui j'ai travaillé ou discuté avec grand plaisir. Pour cela, je souhaite donc remercier Stéphanie dit « la sale gosse », Véronique, Maryse, Maeva, Sylvie, Isabelle, J.F. (merci pour les bières), et Stéphane dit « Camarade » !

Je remercie également tous les membres de l'équipe de l'écloserie : Florence, Christophe, Pascal P, Pascal S, J.C et Raphaël pour leur bonne humeur, leur assistance technique et coups de main, pour assurer le bon déroulement des manips nécessaires à mes travaux. Le nursage des familles F2 a été assuré par Max Nourry, Mathias Papin et Joël Haure, de la station Ifremer de Bouin, c'est pourquoi, je les en remercie.

Cette vague de remerciements se continue par ceux que j'adresse à Martine, Maryse, Florence et Véronique pour l'aide qu'elles m'ont apporté dans l'organisation de mes missions, l'envoi de courriers et de colis, mes demandes bibliographiques et la gestion de toutes les commandes, tant attendues !

Pendant plus de trois ans, j'ai ramé dans une barque que j'ai joyeusement partagé avec les autres doctorants du laboratoire. A cet égard, je remercie Delphine, Sophie, Yann, Julien et Benjamin pour ses expressions si caractéristiques du sud-ouest, qu'il m'a appris (Tête de frangipane !). Je leur souhaite bon courage pour la suite... accrochez vous !

Entre 2003 et 2004, j'ai travaillé en tant que stagiaire sous la direction de deux autres doctorants, deux « anciens » qui m'ont précédé et dont je me suis inspiré : il s'agit de Lionel et Nicolas, que je remercie de m'avoir offert ces stages et pour leurs nombreux conseils !

Au cours de mon doctorat, j'ai quitté épisodiquement ce microcosme pour voyager. J'ai donc eu l'opportunité de visiter l'île de Gran Canaria (Iles Canaris), San Diego (USA) à deux reprises pour participer au colloque du PAG, Birmingham (Royaume-Uni) et enfin, Edinbourg, en Ecosse. J'ai eu un réel coup de cœur pour cette ville car son architecture et son ambiance sont exceptionnelles ! J'espère avoir l'occasion d'y retourner un jour, notamment au moment d'Hogmanay... J'y ai résidé un peu plus d'un mois pour travailler au Roslin Institute avec Chris Haley et DJ de Koning, qui m'ont initié aux joies de la cartographie génétique et de la détection de QTL. Je mesure la chance que j'ai eu de travailler avec eux. Tous deux ont été très prompt à m'aider et à répondre à mes questions les plus farfelues sur la détection de QTL, pour tirer le meilleur de mes données.

Je remercie également très vivement Nicolas Bierne qui m'a beaucoup appris en évolution moléculaire et qui m'a aidé à rédiger l'un des articles présenté dans ma thèse.

Je souhaite remercier tout particulièrement Edwige Quillet, qui a assisté à tous mes comités de thèse. Edwige a toujours su m'encourager et m'a aidé à y voir un peu plus clair dans ce que je faisais avec la tête dans le guidon !

Depuis 2004, j'ai été employé à deux reprises en CDD par la société Genindexe située à La Rochelle. A cet égard, je remercie beaucoup Corinne et Sophie ! Je vous souhaite bon courage !

Je remercie également Brigitte Hudelaine et Jenifer De la Corté-Gomez de l'école doctorale de l'université de La Rochelle pour leur aide.

Je terminerai tous ces remerciements par ceux qui me sont les plus proches, malgré la distance. Ils sont tous venus en Charente Maritime pour y passer quelques jours de vacances et me rendre visite. Je remercie vivement toute ma famille en commençant par ma sœur, mon frère, mes oncles et tantes mais aussi mes beaux-parents et bien évidemment mes parents pour leur soutien moral et financier !

Enfin, je ne sais comment remercier celle qui m'a accompagné tout au long de ces années avec fidélité, qui a ensoleillé mes journées les plus sombres, qui a toujours cru en moi, qui m'a surtout supporté et que j'aime tant. Merci à toi, Adeline.

*"Science sans conscience,
n'est que ruine de l'âme"*

François Rabelais

Sommaire

Chapitre I : Introduction générale

1.1 Biologie de l'espèce	19
1.1.1 Cycle de vie.....	19
1.1.2 Anatomie à l'âge adulte.....	21
1.1.3 Systématique	22
1.2 Elevage et production aquacole.....	24
1.3 Connaissances sur le génome de l'huître creuse	29
1.3.1 Taille et composition du génome	30
1.3.2 Polymorphisme génétique	30
1.3.3 Marqueurs moléculaires	31
1.3.4 Fardeau génétique	32
1.3.5 Cartographies génétiques	33
1.3.6 Cartographie cytogénétique.....	33
1.4 Contexte de l'étude	34
1.4.1 Les caractères d'intérêt pour l'amélioration chez les huîtres	34
1.4.2 Le caractère de survie aux mortalités estivales	35
1.4.3 Etude des bases génétiques de la mortalité estivale du naissain et des possibilités d'amélioration de la survie par sélection dans le cadre du programme Morest	36
1.4.3.1 Sélection en première génération (G_0 - G_1).....	36
1.4.3.2 Seconde génération : sélection divergente pour une faible ou une forte survie du naissain sur estran en première période estivale (G_1 - G_2)	38
1.4.3.3 Troisième génération : répétabilité de la réponse à sélection divergente pour une faible ou une forte survie du naissain en première période estivale (G_2 - G_3)	38
1.4.3.4 Etude des performances de survie du naissain au laboratoire et corrélations avec les données obtenues sur estran.	40
1.5 Objectifs généraux du travail de thèse	44

Chapitre II : Caractérisation phénotypique de familles en période estivale

2.1 Production de familles en ségrégation	49
2.2 Implication d'organismes pathogènes dans les phénomènes de mortalité estivale chez l'huître creuse.....	51
2.3 Production du matériel biologique	53
2.3.1 Choix des géniteurs	53
2.3.2 Plan de croisement	55
2.3.3 Obtention des gamètes et fécondation.....	55
2.3.4 Elevage larvaire	56
2.3.5 Nurserie	56
2.3.6 Suivi de mortalité	57
2.3.7 Extraction d'ADN	59
2.3.8 Protocole de quantification de la charge virale par PCR-Q	59
2.3.9 Outils statistiques	63
2.4 Résultats	63
2.4.1 Elevage larvaire.....	63
2.4.2 Taux de fixation et micronursage.....	64

2.4.3 Phase de nursage	64
2.4.4 Suivi de mortalité	66
2.4.5 Extraction d'ADN	66
2.4.6 Mesure de la charge virale individuelle	67
2.5 Discussion	70
2.6 Perspectives	73

Chapitre III : Développement de marqueurs moléculaires et génotypage

3.1 Polymorphisme et marqueur moléculaire	77
3.2 Les marqueurs Microsatellites	79
3.3 Revue des méthodes de génotypage des marqueurs moléculaires SNP81	
3.3.1 Caractéristiques des marqueurs SNP	81
3.3.2 La détection de SNP	84
3.3.2.1 La détection de SNP par approche « In Silico »	84
3.3.2.2 Le séquençage direct	85
3.3.2.3 Le polymorphisme de conformation de l'ADN	85
3.3.2.4 Le « Tilling »	86
3.3.3 Le génotypage des SNP	86
3.3.3.1 Restriction enzymatique	87
3.3.3.2 L'extension d'amorce	87
3.3.3.3 Le Pyroséquençage	88
3.3.3.4 La mesure du point de fusion par PCR en temps réel (rt-PCR)	88
3.3.3.5 Le SNPlex	91
3.3.4 Applications des SNP	93
3.3.4.1 Cartographie Génétique	93
3.3.4.2 Ancrage d'une cartographie physique	94
3.4 SNP vs Microsatellites	94
3.4.1 Traçabilité d'individus ou d'un produit	94
3.4.2 Assignation de Parenté	94
3.4.3 Cartographie génétique et recherche fine de QTLs	95
3.5 Méthodes de développement et génotypage des marqueurs	96
3.5.1 Extraction de l'ADN	96
3.5.2 Développement de microsatellites In silico	96
3.5.3 Génotypage des marqueurs microsatellites publiés	97
3.5.4 Protocole expérimental	97
3.5.5 La catégorisation des groupes d'allèles ou « Binning »	98
3.5.6 Développement de marqueur de type SNP	99
3.5.7 Analyse du biais d'usage du code génétique chez <i>C. gigas</i>	103
3.6 Critères de choix des loci polymorphes et leur génotypage chez les individus F2	104
3.7 Résultats	105
3.7.1 Microsatellites	105
3.7.2 Développement de marqueurs microsatellites par approche « In Silico »	109
3.7.3 Développement de marqueurs moléculaires de type SNP	115
3.7.4 Génotypage des marqueurs SNP par SNPlex chez les individus de la F2	127
3.8 Discussion	127
3.8.1 Développement de marqueurs SNP	127
3.8.2 Génotypage des marqueurs microsatellites	128
3.8.3 Biais d'usage du code génétique	129
3.9 Perspectives de développement de marqueurs moléculaires chez <i>C. gigas</i>	130

Chapitre IV : Construction d'une cartographie génétique

4.1. Définition et Principe	135
4.1.1 Les tests de liaison entre les loci	136
4.1.2 Le calcul des distances génétiques	136
4.1.3 Test de conformité aux lois Mendéliennes (ou ségrégation).....	140
4.2 Cartographies génétiques établies chez les principales espèces d'intérêt aquacole.....	142
4.3 Matériel et Méthodes.....	149
4.3.1 Matériel biologique et marqueurs moléculaires	149
4.3.2 Choix du logiciel de cartographie	149
4.3.3 Principe d'utilisation du logiciel Crimap	150
4.3.4 Application au jeu de données chez <i>C.gigas</i>	151
4.3.5 Représentation graphiques des cartes.....	152
4.3.6 Comparaison avec la carte génétique précédente établie par Hubert et Hedgecock chez <i>Crassostrea gigas</i>	153
4.4 Résultats	153
4.4.1 Distorsion de ségrégation	153
4.4.2 Carte génétique consensus.....	153
4.4.3 Carte consensus sexe spécifique	157
4.4.4 Comparaison avec la carte génétique développée par Hubert et Hedgecock.....	161
4.5 Discussion	163
4.5.1 Facteurs influençant la construction des cartes génétiques.....	163
4.5.1.1 Distorsion de ségrégation	163
4.5.1.2 Qualité de l'ADN extrait	165
4.5.1.3 Taille de la Population/Pourcentage de données manquantes/Erreurs de génotypes.....	165
4.5.2 Comparaison de cartes au sein de l'espèce <i>Crassostrea gigas</i>	166
4.5.3 Recombinaison différentielle entre mâles et femelles.....	167
4.6 Perspectives de cartographie génétique chez <i>Crassostrea gigas</i>	168
4.6.1 Poursuite du développement de marqueurs moléculaires	168
4.6.2 Intégration des différents types de cartes	169

Chapitre V : Détection de QTL

5.1 Principe de la détection de QTL.....	175
5.2 Etapes des analyses QTL	176
5.3 Modèles de cartographie de QTL.....	177
5.4 Paramètres des analyses QTL	179
5.4.1 Nombre et type de marqueurs	179
5.4.2 Distribution des marqueurs	181
5.4.3 Echantillonnage et schéma expérimental	182
5.4.4 Le génotypage sélectif.....	182
5.4.5 Seuil de signification	183
5.4.6 Estimation de l'intervalle de confiance de la localisation des QTL par « Bootstrapping »	184
5.4.7 Part de variance expliquée.....	184
5.5 Recherche de QTLs chez les espèces d'intérêt aquacole	186
5.6 Matériel et méthodes	187
5.6.1 Simulations de la puissance de détection des QTL	187
5.6.2 Jeu de données disponible pour les analyses QTL.....	187

5.6.3 Présentation du logiciel QTL express	187
5.6.4 Module F2	188
5.6.5 Module Half-Sib	189
5.6.6 Module d'analyse en intra-famille – module « Tree »	190
5.7 Résultats	191
5.7.1 Module « F2 »	191
5.7.2 Module « Half sib ».....	199
5.7.3 Module « Large single full-Sib family analysis- Tree »	199
5.8 Discussion	201
5.8.1 Nombre de QTL identifié et part de variance expliquée.....	201
5.8.2. Ségrégation différentielle entre les familles F2.....	203
5.8.3. Relations entre QTL de survie et de charge virale.....	204
5.9 Perspectives.....	205

Chapitre VI : Conclusion et synthèse générale

6.1 Poursuite des travaux de cartographies	211
6.2 Détection de QTL.....	212
6.3 Interactions hôte-pathogène	213
6.4 Applications pour la filière conchylicole	214
Bibliographie.....	219
Liste des figures et des tableaux.....	253
Annexes.....	261

CHAPITRE I

INTRODUCTION GENERALE

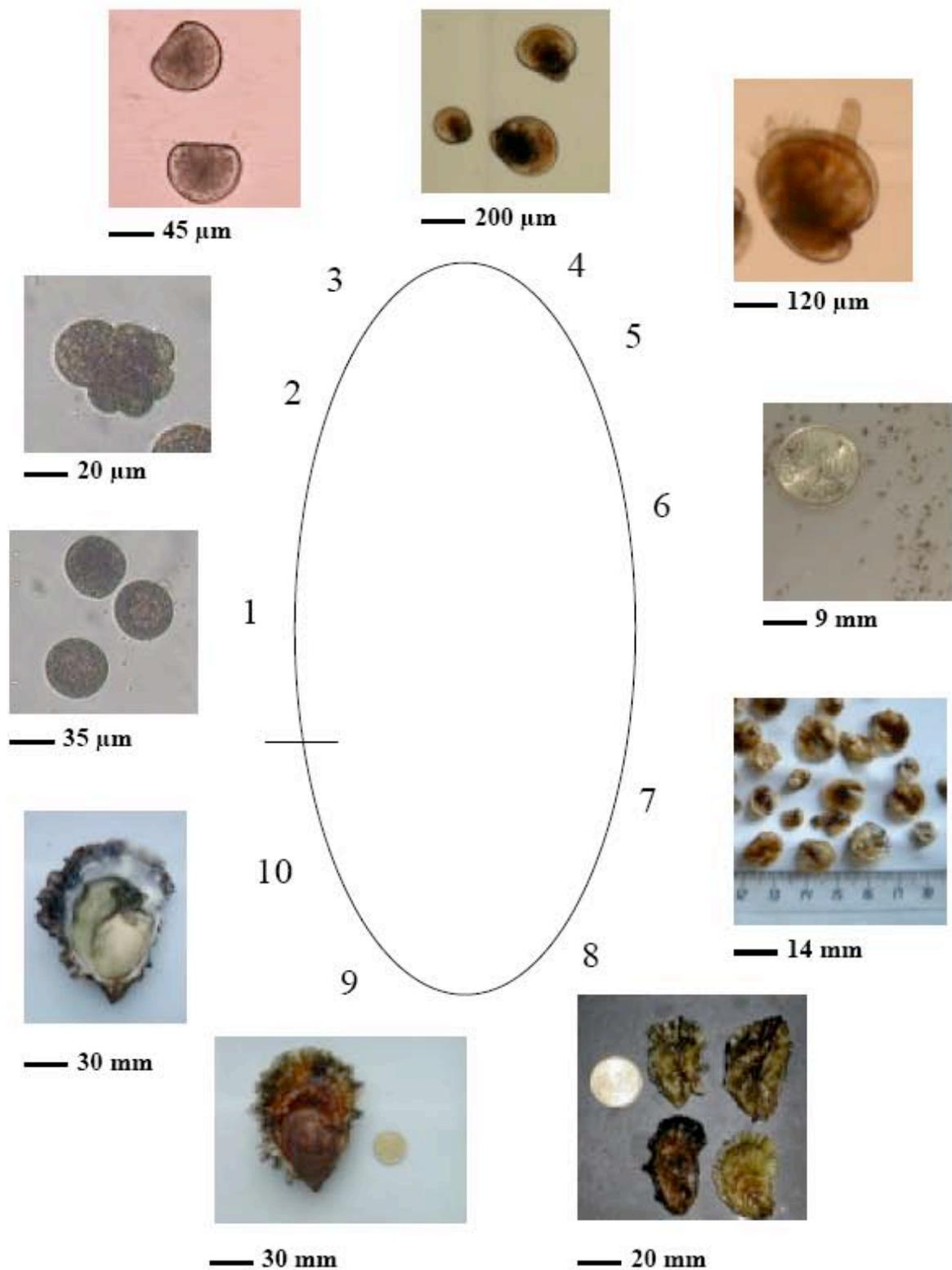


Figure 1 : Cycle biologique de l’huître creuse *Crassostrea gigas*. 1-Fécondation : ovocytes en présence de spermatozoïdes. 2-Embryon stade morula (2-3 h). 3-Larves D (24 h). 4-Larves véligères (14 jours). 5-Larve pédivéligère (18 jours). 6- Naissains post-fixation (1 mois). 7- Naissains (2 mois). 8-Naissains (6 mois). 9-Adulte (10 mois). 10-Géniteur mature (10 mois) (Dégremont, 2003).

Ce chapitre a pour objectif de passer en revue les connaissances sur la biologie, le génome et la production nationale et internationale de l'huître creuse *Crassostrea gigas* et de recadrer ce travail de thèse dans le contexte de la sélection génétique au cours du projet Morest.

1.1 Biologie de l'espèce

1.1.1 Cycle de vie

L'huître creuse est une espèce ovipare qui se caractérise par une forte fécondité (Walne, 1974). Le déterminisme sexuel de cette espèce reste mal connu. Cependant, l'huître creuse est hermaphrodite à tendance protandre ; c'est à dire que les individus sont généralement mâle au cours des premiers cycles de reproduction puis se comportent comme femelles (Buroker, 1983). Son cycle de vie se déroule en deux phases principales (figure 1) : une phase larvaire libre ou pélagique suivi d'une phase benthique ou fixée. Entre le mois de mars et le mois d'octobre, les individus adultes, âgés de 1 an et plus et d'une taille d'au moins de 2 cm de longueur, entament leur gamétogenèse pour atteindre la maturité sexuelle à l'approche de la période estivale. La gamétogenèse est influencée par de nombreux facteurs internes et externes tels que l'état physiologique et immunitaire de l'animal, l'abondance d'éléments nutritifs et la température du milieu (Gérard et al., 1995); (Gouletquer, 1997); (Fabioux, 2004). La fécondation est externe et les gamètes sont expulsés dans le milieu. Les premières divisions cellulaires interviennent très rapidement après fécondation, le stade morula (16 cellules) étant atteint après 3 h. Après 24 h de développement, la larve trocophore mesure environ 70 μm , possède une coquille et un velum qui lui sert d'organe de nutrition et de locomotion : c'est le stade de la larve «D» en référence à la forme de la coquille. La larve va poursuivre son développement pendant quinze à vingt cinq jours environ selon la température de l'eau et de l'abondance en éléments nutritifs. Elle atteint généralement une taille de 300 à 380 μm de diamètre avant de se métamorphoser. Un organe sensoriel, « l'œil », apparaît au moment de cette transformation. Cet organe prend l'apparence d'un point noir très réfringent au milieu de la coquille de la larve. Parallèlement à cet œil, la larve a également développé un « pied » (stade pédivéligère). Ce dernier lui permet de rechercher un substrat solide propice à sa fixation. La larve se métamorphose alors et passe d'une vie pélagique à une vie benthique : le stade juvénile est alors atteint. L'individu ressemble à un adulte à l'exception de sa petite taille. Ensuite, l'animal va croître puis entrer en gamétogenèse, dès sa première année de vie si les conditions sont favorables.

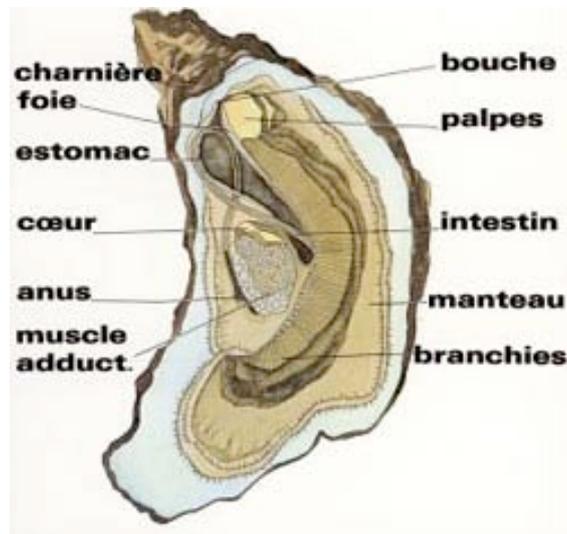


Figure 2 : Anatomie à l'âge adulte de l'huître creuse *Crassostrea gigas*

1.1.2 Anatomie à l'âge adulte

Une vue générale de l'anatomie de l'huître creuse est représentée en figure 2. L'huître vie protégée par sa coquille dont l'ouverture et la fermeture est assurée par le muscle adducteur (figure 2). Cette coquille est composée de deux valves (gauche et droite) asymétrique reliées par une charnière. L'une des deux valves est creusée et accueille ainsi la masse viscérale de l'animal. Cette caractéristique morphologique donne son nom usuel à l'huître creuse (Hughes, 2002). La masse viscérale de l'animal est composée de plusieurs organes. Tout d'abord, le manteau, composé de deux lobes, renferme la cavité palléale et assure diverses fonctions. C'est un organe sensoriel mais surtout l'organe qui assure la filtration des éléments nutritifs pour l'alimentation de l'animal. De plus, il permet la croissance de la coquille et la synthèse de nacre qui recouvre l'intérieur de cette coquille. La cavité palléale comprend les orifices anal et rénaux ainsi que les branchies.

L'huître creuse possède une paire de branchies composées de quatre feuillets branchiaux, constituées de rangées longitudinales de filaments. Cet organe a deux fonctions : tout d'abord, les branchies retiennent les particules nutritives en suspension (fonction de nutrition) ; ensuite, elles permettent l'absorption de l'oxygène dissous dans l'eau (fonction de respiration Gouletquer et al., 1999). L'œsophage est cilié et donne sur l'estomac qui contient un stylet cristallin. L'estomac se prolonge par l'intestin et le rectum qui longe le muscle adducteur et se termine par l'orifice anal. L'organe reproducteur ou gonade est diffus et se développe autour de la glande digestive. Sa taille varie considérablement en fonction de la période de l'année. En effet, lorsque la maturité sexuelle est atteinte, généralement au mois de juin, celle-ci peut atteindre 70% de la masse sèche de l'animal sans différenciation de sexe (Gouletquer, 2005). Le cœur, formé de deux oreillettes et de deux ventricules, se situe dans la cavité péricardique à proximité du muscle adducteur. Enfin, le système nerveux est réduit, composé de ganglions cérébroïdes et viscéraux.

1.1.3 Systématique

Le nom usuel de *Crassostrea gigas* est l'huître japonaise ou huître creuse du Pacifique. Il s'agit d'un mollusque bivalve marin dont la position systématique précise est la suivante, selon les critères de Grasset établis en 1960 :

Règne Animal
Embranchement des *Mollusca*
Classe des *Bivalva*
Ordre des *Filibranchia*
Sous ordre *Pteromorphia*
Super famille des *Ostreoidea*
Famille des *Ostreidae*
Genre *Crassostrea*
Espèce *gigas*

Il est à noter que la taxonomie chez les huîtres est parfois ambiguë. La grande plasticité morphologique des coquilles peut facilement porter à confusion et la plupart du temps, les caractères de la coquille sont utilisés pour l'identification. Ranson (1948) est malgré tout parvenu à distinguer dans la famille des *Ostreidae* trois genres (*Crassostrea*, *Ostrea* et *Pycnodonta* d'après les caractéristiques de la prodissoconque (coquille larvaire). Harry (1985), toujours selon des critères morphologiques, ajoute trois sous-familles à la famille des *Ostreoidea*, à savoir *Lophinae*, *Ostreinae*, et *Crassostreinae* à laquelle le genre *Crassostrea* appartient.

Le genre *Crassostrea* regroupe près de 6 espèces différentes dont la répartition géographique est large puisqu'elle s'étend de l'Asie (*C. gigas*, *C. ariakensis*, *C. sikamea*); à l'Afrique (*C. gasar*) et l'Amérique (*C. virginica*, *C. rhizophorea*). Cependant, la notion d'espèce est relativement complexe et parfois mal élucidée (Gaffney et al., 1993), plus particulièrement entre les espèces *C. gigas* et *angulata*. Différentes études, morphologiques (Ranson, 1960; Menzel, 1974), expériences d'hybridation (Huvet et al., 2001; Huvet et al., 2002), analyse caryotypique (Leitão, 1999), marqueurs allozymiques (Buroker et al., 1979; Mattiucci et al., 1983) et ADN satellite (Lopez-Flores et al., 2004) considèrent les deux taxa comme une même espèce. D'autres données différencient ces deux espèces (caractéristiques écophysiologicals ; Haure et al., 2003); par analyse de caryotypes (Leitao et al., 2004) appuyées par des analyses d'ADN mitochondrial (Boudry et al., 1998; O'Foighil et al., 1998; Boudry et al., 2003a). Le séquençage de l'huître creuse *C. gigas* (Hedgecock et al., 2005) est un argument supplémentaire permettant de différencier de manière certaine, les différentes espèces du genre *Crassostrea* .



Figure 3 : Principaux pays producteur de l’huître creuse *C. gigas* (FAO Fishery Statistics, 2002)

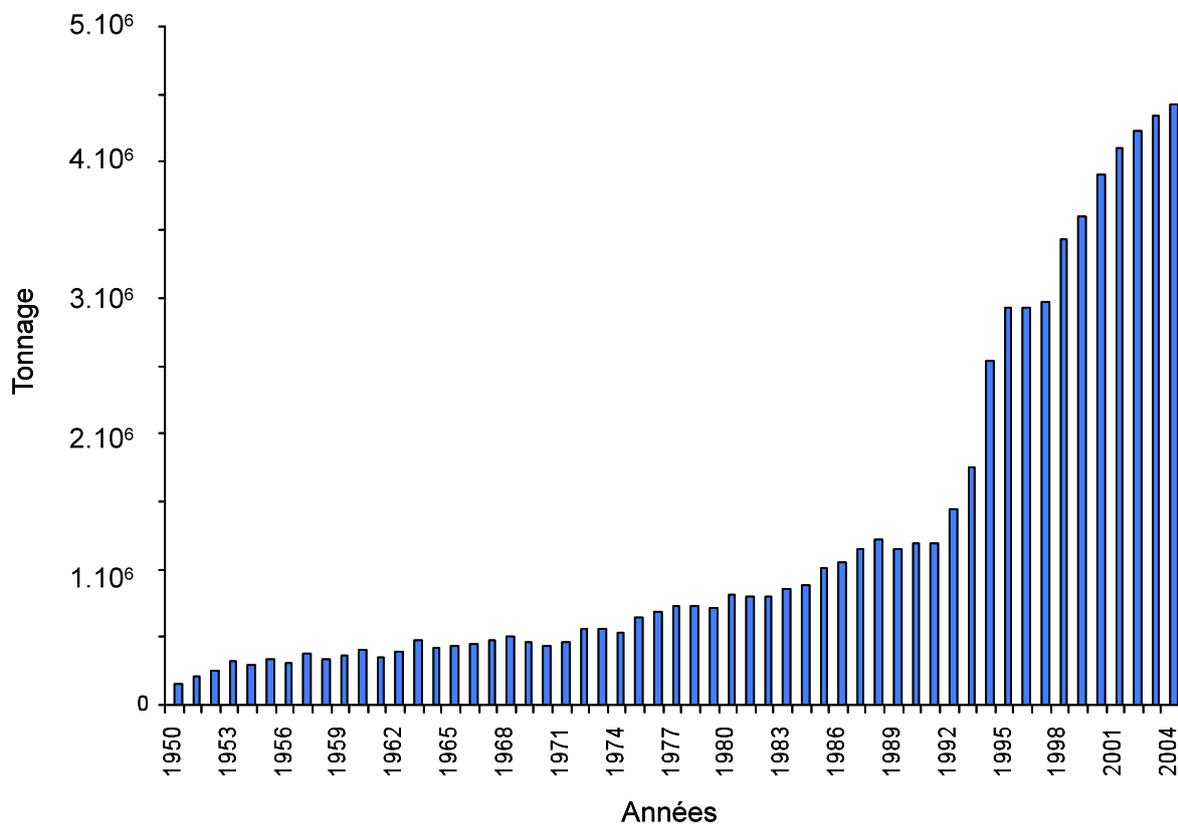
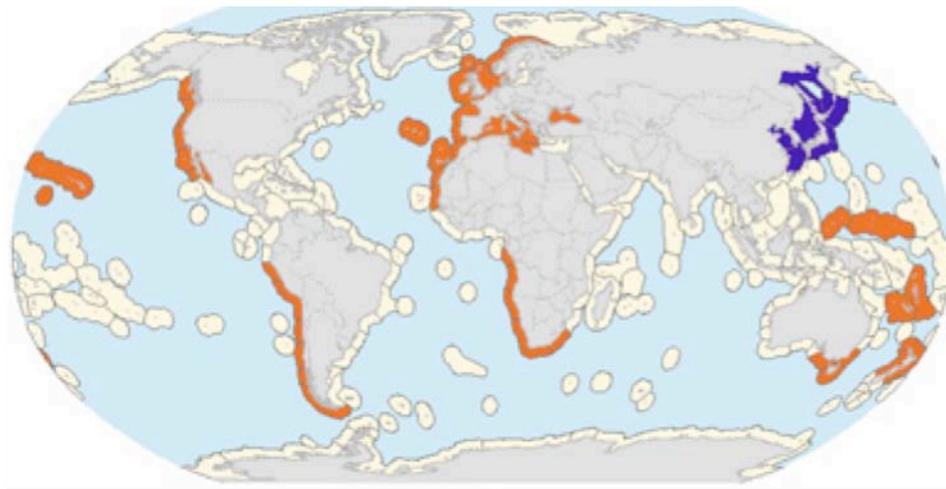


Figure 4 : Production annuelle mondiale de l’huître creuse (en tonnes, FAO, 2006).

1.2 Elevage et production aquacole

Originnaire d'Asie, et plus précisément du Japon, l'huître creuse a été introduite en France en 1966 (Grizel et al., 1991). Celle-ci coïncide avec l'épisode de mortalité massive de l'huître portugaise *Crassostrea angulata*. Ces mortalités sont attribuées à une infection présumée d'origine virale (iridovirus) plus connu sous le nom de maladie de la branchie (Comps et al., 1976). Sa répartition géographique naturelle s'étend de l'Asie du Nord Est, à l'Europe depuis le Portugal à la Norvège et en Amérique du Nord du Sud est de l'Alaska à la Californie (figure 5). Cette espèce a également été introduite en Afrique, dans la région Australo-Pacifique et en Amérique du Nord et du Sud (ICSEMS, 2000; AMCS, 1998; Hopkins, 2001; Boyd, 2003). Depuis cette époque, l'huître du Pacifique est l'espèce dominante en France. En 2006, la production globale de cette espèce était estimée à 4,5 millions de tonnes (Mt) pour une valeur marchande estimée à environ US\$ 3,7 milliards. Cela représente la production la plus importante en tonnage devant celles des autres poissons ou crustacés. Près de 84% (3.7Mt) de cette production globale est aujourd'hui réalisée sur le littoral chinois. Le reste de la production est majoritairement assurée par le Japon (261.000 tonnes) ; la République de Corée (238.000 tonnes) et la France (115.000 tonnes). Les Etats-Unis (43.000 tonnes), Taiwan (23.000 tonnes) produisent également cette espèce (figure 3). La production mondiale de cette espèce croît de manière régulière (figure 4) : elle est passée de 156.000 tonnes en 1950, à 437.000 tonnes en 1970. Cependant, c'est au cours des années 1990 que la production « explose » pour atteindre 4,5 millions de tonnes en l'an 2006, notamment par l'augmentation de la production chinoise. A l'heure actuelle, la production de l'huître creuse est toujours en augmentation mais elle tend à se ralentir à cause, notamment, de l'urbanisation de la zone côtière et des zones de culture (FAO, 2007).



- Zone native
- Espèce présente
- Présence non décrite

Figure 5 : Répartition naturelle mondiale de l’huître creuse *C. gigas* et zones d’introduction

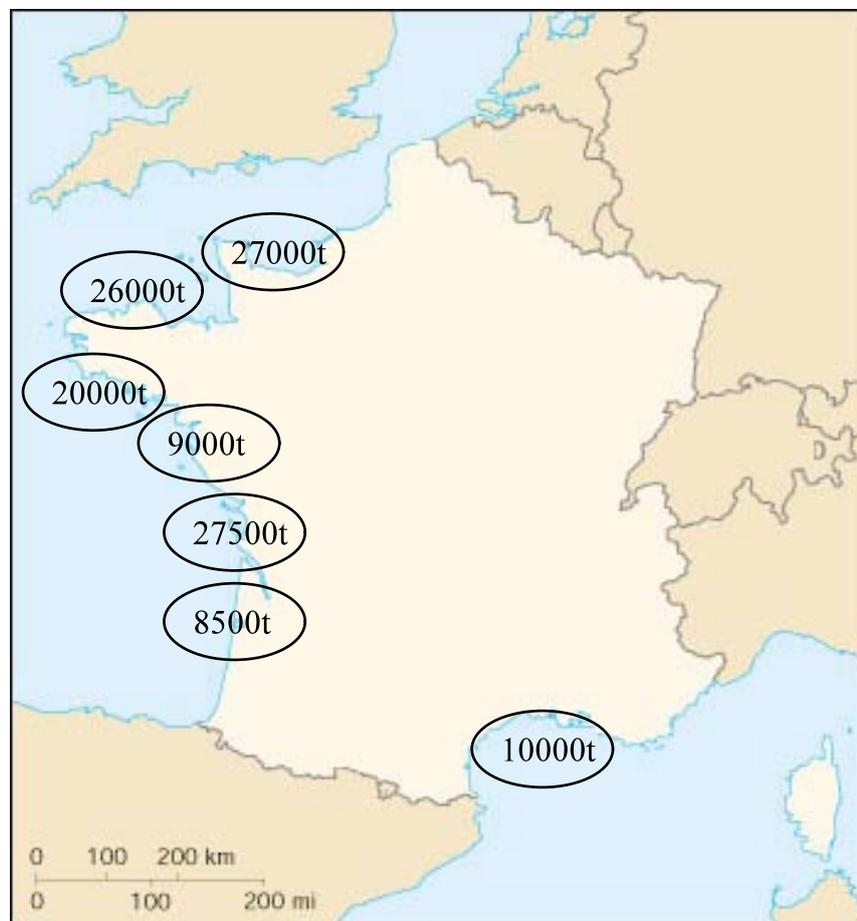


Figure 6 : Répartition de la production de l’huître creuse sur les sites d’élevage français (données du Centre National de la Conchyliculture, 2006).

La France occupe aujourd'hui le quatrième rang mondial et le premier européen pour la production de l'huître japonaise. En 2001, cette production reposait à 98% sur l'exploitation de *C.gigas* et cette production était issue à 90% du captage naturel de naissain. Ce captage était principalement réalisé sur le bassin d'Arcachon et celui de Marennes-Oléron.

La production française est localisée sur une plus grande zone géographique (figure 6) qui s'étale sur la façade maritime méditerranéenne, atlantique, bretonne et normande. En effet, le cycle d'élevage que pratique les professionnels de l'ostréiculture se déroule communément de la manière suivante : au cours de la période estivale, les ostréiculteurs placent sur l'estran des collecteurs de nature diverses (tuiles chaulées ; coupelles plastiques, coquilles d'huîtres) afin de permettre aux larves de trouver un substrat propice à leur fixation. Puis, après une période de croissance de 6 à 18 mois, le naissain est détaché, c'est à dire qu'il est détaché manuellement de son support pour être mis en poche. Les poches sont alors placées sur des tables ostréicoles dont la hauteur varie entre 15 et 70 cm de hauteur. Une nouvelle période de croissance commence alors. Pour les ostréiculteurs, le travail consiste à retourner régulièrement les poches afin d'éliminer le film algal qui se développe sur les poches, puis de changer la maille des poches et enfin de les dédoubler en fonction de la croissance des animaux. Ceux-ci atteignent une taille commerciale dès 40 grammes. Les huîtres peuvent alors être disposées en claires ostréicoles pour affinage afin de répondre à des critères de qualité dont dépendent les appellations commerciales. D'autres pratiques culturelles consistent à semer les huîtres 'à plat' sur estran ou en eaux profondes.

Cependant, les méthodes de production d'huîtres ne reposent plus uniquement sur le captage naturel qui apparaît moins régulier (<http://www.theveliger.org/>). En effet, la production de naissain issue d'écloserie commerciales est en pleine expansion. Les techniques de reproduction artificielle ont été améliorées depuis les premiers travaux (Brooks, 1879). Désormais, ces techniques sont bien maîtrisées pour assurer la reproduction en masse (Helm et al., 2006) dans des conditions différentes de celles du milieu naturel (température supérieure à 25°C, apport trophique présumé optimal). Ces conditions intensives d'élevage induisent des pressions de sélection (perte de diversité génétique et d'effectif efficace, Hedgecock et al., 1992) entraînant un phénomène de sélection-domestication dès les stades précoces de développement (Taris, 2005).

La part de naissain produit en écloseries occupait environ 10% du naissain produit en 1998 (Robert et al., 1999) En 2004, 50% du naissain produit est issu d'écloserie (Leroux, 2004). L'optimisation des performances de *C. gigas* s'est orientée vers la manipulation du

niveau de ploïdie. Ces recherches ont permis d'obtenir des huîtres creuses triploïdes stériles, à partir du croisement d'huîtres males tétraploïdes par des femelles diploïdes, dont la réduction gonadique permet un meilleur taux de croissance et une commercialisation étalée sur toute l'année. Il a été démontré que l'huître triploïde présentait des performances de croissance constantes toute l'année (Allen et al., 1986), ainsi qu'une qualité de chair et une résistance aux pathogènes tel que *Vibrio parahaemolyticus* beaucoup plus développées que l'huître diploïde commune (Peyre et al., 1999). L'intérêt aquacole de l'utilisation d'individus triploïdes a conduit, ces dernières années, au développement de cette production en France (Nell, 2002).

L'huître creuse *Crassostrea gigas* introduite massivement en France dans les années soixante dix constitue désormais 70 % de la conchyliculture française. L'espèce se trouve à la limite des conditions naturelles de sa reproduction. En conséquence, le risque écologique de sa dispersion n'avait pas été initialement envisagé. Toutefois, après s'être reproduit « naturellement » dans les bassins conchylicoles, un processus de colonisation a été observé depuis plus de 10 ans jusqu'en Bretagne et Normandie et plus largement en Belgique, aux Pays bas et au Royaume-Uni. L'espèce présente aujourd'hui des caractéristiques proches d'une espèce invasive. Au niveau mondial, *C. gigas* a été introduite dans 70 pays dans le cadre d'activités aquacoles. Cette espèce fait désormais l'objet d'études d'impact sur l'environnement côtier de part sa prolifération car ce phénomène entraîne des conséquences écologiques sur les populations naturelles mais aussi socio-économiques (ex : impact sur le rendement de production de moules aux Pays-Bas).

En France, étant donné l'importance du phénomène, des opérations massives de gestion des bancs naturels ont été effectuées sur les principaux sites de captage (Arcachon et Marennes) pour diminuer la compétition trophique et augmenter la productivité des élevages. Aujourd'hui, de nombreux habitats marins reconnus dans le cadre de la Directive Habitat sont touchés par cette prolifération alors qu'ils sont déjà soumis à de fortes pressions et des gestionnaires ont pris des mesures d'éradication qui restent toutefois très locales.

Malgré l'ampleur du phénomène, sa rapidité a fait qu'aucune étude scientifique n'a pour le moment été réalisée et a fortiori aucune mesure de gestion n'a été proposée pour gérer cette colonisation déjà présente dans certains sites Natura 2000.

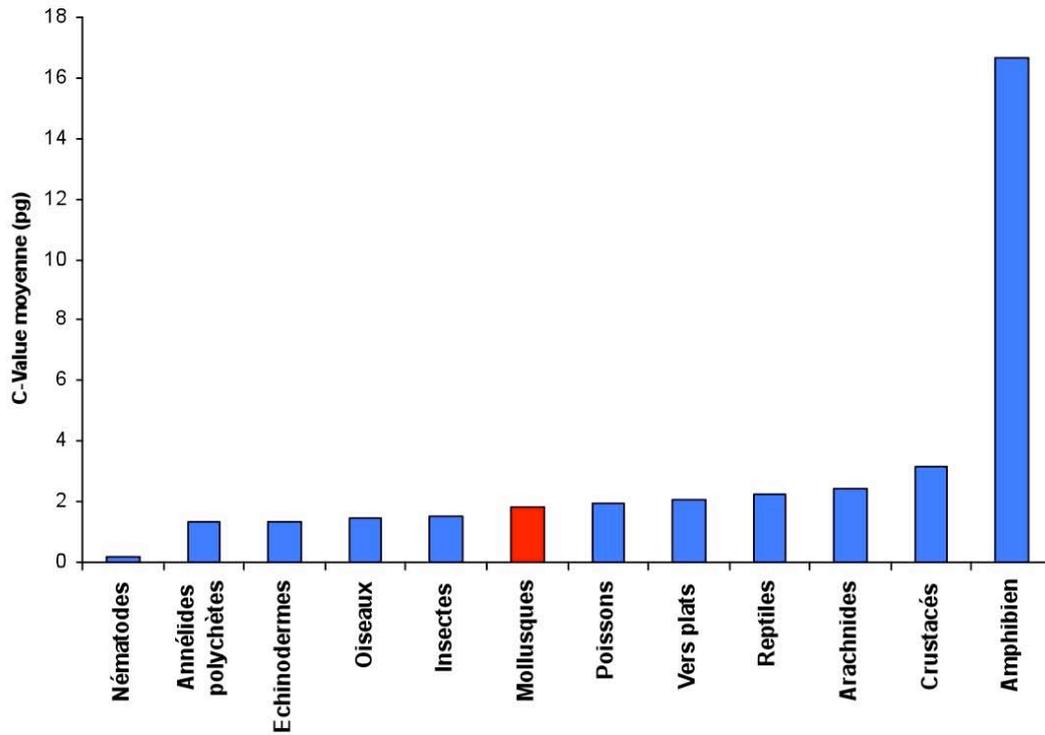


Figure 7 : Taille de génomes chez les vertébrés et invertébrés (pg) (Gregory, 2005)

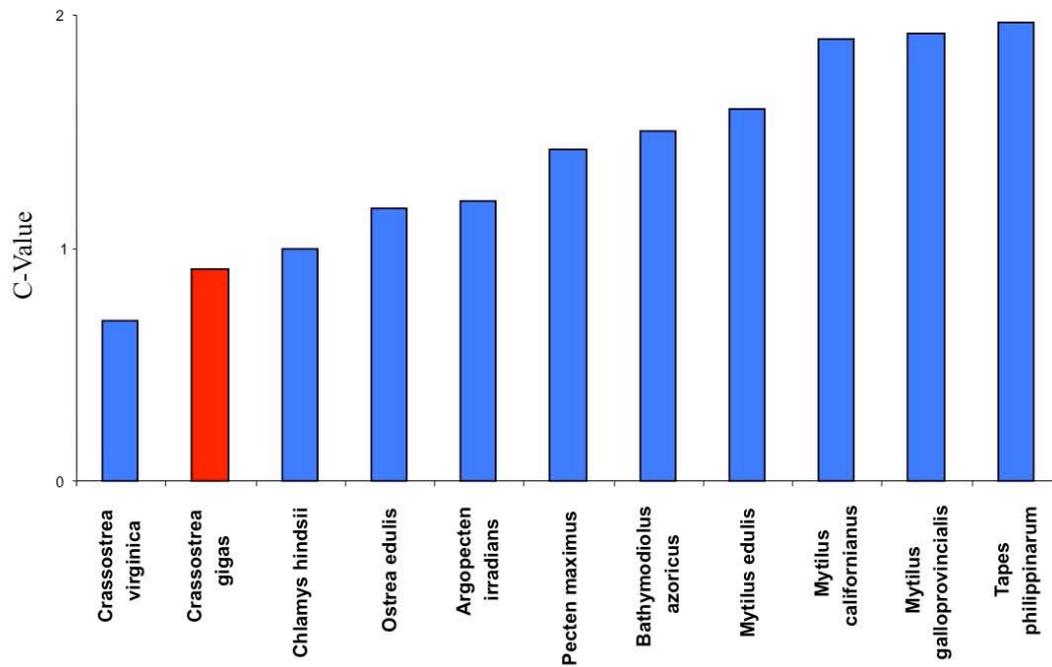


Figure 8 : Estimation de la taille de génomes chez quelques bivalves (pg) (Gregory, 2005)

1.3 Connaissances sur le génome et la génétique de l'huître de l'huître creuse

La taille du génome chez les mollusques bivalves est comprise entre 0,65 pg et 5,4 pg alors que, sur l'ensemble du règne animal, cet intervalle s'étend de 0,03 pg chez le nématode *Meloidogyne graminicola* à 133 pg chez le Protoptère d'Ethiopie, *Protopterus aethiopicus* (Gregory, 2005). Cet intervalle se trouve dans la moyenne de la taille des génomes d'espèces modèles telles que certains poissons ou insectes (figures 7 et 8). Des données ont été obtenues sur le génome de l'huître creuse *Crassostrea gigas* car cette espèce représente d'importants intérêts économiques (FAO, 2007).

Les EST (Expressed Sequence Tags) sont classiquement des séquences d'ADN obtenues à partir d'une librairie d'ADN codant (ou ADNc). Les séquences EST sont très utiles car elles apportent des informations sur la part du génome exprimé chez l'organisme et elles sont également à la base de nombreuses applications en génomique telles que le développement de microarray. Environ 22000 ESTs sont disponibles chez *C. gigas* et sont issus, par exemple, de challenge en hybridation soustractive suppressive menés sur différents types cellulaires (Tableau 1 ; Huvet et al., 2004; Tanguy et al., 2008). De nouveaux EST ont été récemment développés au sein du programme de recherche « Marine Genomics Europe » (<http://www.marinegenomics.org>) et déposés sur GenBank enrichissant considérablement les bases de données. Dans un futur proche, plus de 300.000 EST, issus d'un effort de séquençage mené aux Etats-Unis, viendront enrichir les bases de données.

Tableau 1 : Principaux EST développés chez *C. gigas*.

Auteurs Principaux	Tissu(s) cible(s)	Nombre d'ESTs	Publications associées
J.M. Escoubas / E. Bachère	Hémocytes	1142	(Gueguen et al., 2003)
A. Tanguy / X. Guo	Hémocytes / Manteau	70	(Tanguy et al., 2004)
A. Huvet	Manteau / Gonade	137	(Huvet et al., 2004)
A.H. Kausland	Manteau	409	-
A. Tanguy / D. Moraga	Gonade / Glande digestive	377	(Tanguy et al., 2005)
G.P. Rafferty	Manteau	106	-
Marine Genomics Europe	Divers	>15000	(Tanguy et al., 2008)
Génoscope	Divers	>25000	-

Au début de ce travail (2004), le nombre d'EST était relativement faible par rapport à d'autres espèces modèles, mais le nombre de gènes impliqués dans les mécanismes de défense immunitaire, de mécanisme énergétique ou de réponse aux stress a considérablement augmenté. Enfin, plus de 25000 EST vont être déposés dans le domaine public suite au travail réalisé par le Génoscope.

1.3.1 Taille et Composition et taille du génome

Mc Lean et Whiteley (1973) ont été les premiers à caractériser le génome de l'huître creuse. Sa taille fut estimée à 1,26 pg et sa composition à environ 32,2% en bases G et C. Cependant, Gonzalez-Tizon et collaborateurs (2000) ont ré-estimé la taille haploïde du génome de l'huître creuse, par analyse d'images Feulgen (technique de coloration spécifique de l'ADN), à 0,91 pg, soit une taille de huit cents vingt quatre millions de paires de bases (Mb). Le génome de l'huître est d'une taille relativement faible par rapport à d'autres bivalves tels que la moule bleue *Mytilus edulis* (1,6 pg \Leftrightarrow 1400Mb ; figure 7 ; Gregory, 2005). A l'heure actuelle, aucun génome de bivalve, n'a été complètement séquencé. Depuis 2004, Hedgecock et collaborateurs (2005), associés au sein de l'Oyster Genome Consortium (OGC) ainsi que le génome sequencing center de Pékin (Chine) promeuvent le séquençage du génome de l'huître creuse.

Le caryotype diploïde de l'huître creuse comprend 20 chromosomes ($2n=20$) métacentriques et sub-métacentriques (Ahmed et al., 1967; Leitão, 1999; Leitão, 2001). Le phénomène d'aneuploïdie (caractéristique d'une cellule qui ne possède pas le nombre normal de chromosomes, mais 17, 18, ou 19) a été largement étudié. Une corrélation négative entre l'aneuploïdie somatique et le taux de croissance a été décrite dans la descendance d'huîtres cultivées (Thiriot-Quévieux et al., 1988; Thiriot-Quévieux et al., 1992; Leitao et al., 2001) et dans les populations naturelles (Zouros et al., 1996). De plus, l'hypothèse d'une base génétique dans la détermination de l'aneuploïdie a été émise (Leitao et al., 2001) et il existerait aussi une perte préférentielle de certains chromosomes (dans les paires 1, 5, 9 et 10) dans les cellules aneuploïdes (Leitao et al., 2001). Enfin, l'effet de polluants sur le taux d'aneuploïdie des huîtres creuses a également été mis en évidence (Bouilly et al., 2003; Bouilly et al., 2005; Bouilly et al., 2006 ; Bouilly et al., 2007).

1.3.2 Polymorphisme génétique

La première forme de polymorphisme génétique mis en évidence chez *C. gigas* est le polymorphisme protéique. Chez cette espèce, le taux moyen d'hétérozygotie aux marqueurs allozymiques est supérieur à 20% ce qui est trois à quatre fois plus important que le taux d'hétérozygotie observé chez les mammifères (Buroker et al., 1975; Hedgecock et al., 1990). L'abondance d'allèles nuls lors de l'amplification de marqueurs moléculaires de type microsatellites met également en évidence un polymorphe nucléotidique très important (McGoldrick et al., 2000; Huvet et al., 2000; Launey et Hedgecock, 2001; Reece et al., 2004).

Le polymorphisme des marqueurs moléculaires microsatellites est également important. L'amplification de certains marqueurs microsatellites montrent plus de 40 allèles différents par marqueurs (Huvet et al., 2000). Plus récemment, Curole et Hedgecock (2005) ont mis en évidence un niveau de polymorphisme SNP (Single Nucleotide Polymorphism) très élevé avec 1 SNP toutes les 40 paires de bases et une « indel » (insertion-délétion) toutes les 33 paires de bases.

1.3.3 Marqueurs moléculaires

Plusieurs types de marqueurs moléculaires ont été développés chez l'huître creuse dans le cadre d'études de génétique des populations, d'identification d'espèces, d'études de polymorphisme ou de cartographies génétiques. A l'heure actuelle, les marqueurs moléculaires disponibles sont :

- Allozymes: la vigueur hybride (ou heterosis) a été décrite chez les bivalves pour la première fois en 1978 (Singh et al., 1978). Cette vigueur était positivement corrélée au niveau d'hétérozygotie de marqueur allozymes chez des populations naturelles. Les causes de cette association entre marqueurs et l'hétérosis ont été largement débattu et opposèrent deux hypothèses : la superdominance « directe » ou « associative » (Zouros et al., 1994; Britten, 1996; David, 1998; Bierne et al., 2000).
- SSCP (single Strand Conformation Polymorphism) de gènes mitochondriaux ou nucléaires: ces marqueurs ont été utilisés pour différentes application telles que l'étude du succès reproducteur (Li et al., 1998); du polymorphisme de gènes impliqués dans des mécanismes de défense (Tanguy et al., 2002; David et al., 2007).
- RAPD (Random Amplification of DNA Polymorphism) : ces marqueurs, dont l'utilisation est relativement aisée, ont été développés afin d'identifier deux populations d'huîtres provenant du Japon (Hiroshima) et de Corée (Goseong). Au final, 49 marqueurs RAPD particulièrement polymorphes (93% d'individus polymorphes) ont été développés (Aranishi et al., 2004) .
- AFLP (Amplified Fragments Length Polymorphism) : Pour l'élaboration de leur carte génétique, Li et Guo (2004) ont utilisé 17 combinaisons d'amorces afin de développer le maximum de marqueurs AFLP possible. Au total, 1106 pics AFLP

ont tout d'abord été identifiés, mais seulement 34,7% d'entre eux ont été utilisés car ils étaient polymorphes chez les individus issus du croisement backcross utilisé pour cette étude. Cette approche a également été utilisée pour la construction de la carte génétique chez l'huître creuse *Crassostrea virginica*, chez laquelle, 114 marqueurs AFLPs ont été cartographiés (Yu et Li, 2003).

- Microsatellites : les marqueurs microsatellites sont aujourd'hui les marqueurs moléculaires les plus utilisés chez *C. gigas*. Désormais, il existe plus de 300 marqueurs microsatellites déposés dans « GenBank » dont plus d'une centaine a été publiée (Magoulas et al., 1998; Huvet et al., 2000; Li et al., 2003; Sekino et al., 2003; Yamtich et al., 2005; Yu et Li, 2007; Wang et al., 2008; Yu et Li, 2008). Ils ont été largement utilisés pour des études de parenté (Taris, 2005) ou de construction de cartographies génétiques (Li et Guo, 2004; Hubert et Hedgecock, 2004; Yamtich et al., 2005) car leur atout majeur est le nombre moyen élevé d'allèles à un locus (>30 ; Huvet et al., 2004 ; Boudry et al., 2002).
- SNP (Single Nucleotide Polymorphism) : A l'heure actuelle et à notre connaissance ; aucun marqueur de type SNP n'est clairement identifié, ni déposé dans le domaine publique en tant que marqueur moléculaire chez *Crassostrea gigas*, mais il existe un intérêt certain pour leur développement en raison du haut niveau de polymorphisme observé facilitant leur identification au sein des séquences ESTs (Curole et Hedgecock, 2005).

1.3.4 Fardeau Génétique

Chez les mollusques bivalves, des phénomènes d'association entre marqueurs et hétérosis et de distorsion de ségrégation, ont été largement décrits au cours des 25 dernières années. Un important fardeau génétique de mutations délétères récessives pourrait expliquer ces deux phénomènes, selon l'hypothèse de dominance de l'hétérosis. Launey et Hedgecock (2001) ont utilisé des lignées dérivées d'une population naturelle d'huître du Pacifique, ils ont comparé les ratios de ségrégation de microsatellites à 6 h après fécondation et à 2-3 mois en familles hybrides F2 ou F3. Les données de ségrégation des marqueurs, lorsque les modèles de sélection liée aux mutations récessives délétères sont extrapolés à l'ensemble du génome, estiment que ces lignées portent chacune de 8 à 14 des mutations récessives hautement délétères. Cette preuve d'un fardeau génétique élevé établit l'huître comme un modèle animal pour la compréhension des causes génétiques (dépression de consanguinité) et physiologiques de l'hétérosis (Hedgecock et al., 2007).

1.3.5 Cartographies génétiques

Deux cartes génétiques ont été développées chez l'huître creuse. La première est basée sur des marqueurs de type AFLP (Li et Guo, 2004); et la seconde sur une centaine de marqueurs microsatellites (Hubert et Hedgecock, 2004; tableau 2). La première carte regroupe 119 marqueurs chez la femelle et 96 chez le mâle réparti en 11 et 10 groupes de liaison, s'étend sur une distance totale de 1258 et 933 cM respectivement chez la femelle et le mâle. La seconde carte, se divise en 12 et 11 groupes de liaison, s'étend sur une longueur de 770 et 616 cM chez la femelle et le mâle et regroupe 88 et 86 marqueurs microsatellites respectivement. En terme de densité, la première carte est relativement plus dense avec une distance moyenne intermarqueurs de 9,5 et 8,8 cM chez la femelle et le mâle alors que pour la seconde carte, cette même distance, est comprise entre 13 et 10 cM selon le sexe considéré. Ces deux cartographies posent les bases nécessaires à la recherche de QTLs d'intérêt chez cette espèce. Ces aspects de cartographie génétique seront développés plus amplement au sein du chapitre III de ce manuscrit de thèse.

Tableau 2 : Principales caractéristiques des deux cartes de liaison développées chez *C.gigas*.

Nombre de marqueurs (Femelle / Mâle)	Nombre de groupes de liaison (Femelle/Mâle)	Taille de la carte (Femelle/Mâle - cM)	Distance intermarqueur (Femelle/Mâle - cM)	% de couverture du génome (Femelle / Mâle)	Publication associée
119 / 96 (AFLP)	11 / 10	1258 / 933	9,5 / 8,8	82 / 81,3	(Li et al., 2004)
88 / 86 (Microsatellites)	12 / 11	770 / 616	13,6 / 10,1	75 / 79	(Hubert et al., 2004)

1.3.6 Cartographie cytogénétique

Au sein du genre *Crassostrea*, des analyses comparatives de caryotype marqués (G-banding) ont été menées chez trois espèces (*C. gigas*, *C. angulata* et *C. virginica*) mettant en évidence des régions communes (chromosome 3) et des régions caractéristiques de chacune des espèces, fournissant un argument supplémentaire à la divergence génétique entre *C. gigas* et *C. angulata* (Leitao, 2004; Leitão, 1999). De plus, la cartographie cytogénétique de gènes ribosomiaux (sous unité 5S) par hybridation *in situ* a été réalisée (Wang et al., 2001). La séquence cible a été identifiée sur les paires chromosomiques 1, 2, 4, 5, 7, 8 et 10. A l'exception de ces travaux, peu de données sont disponibles. Chez les espèces proches *C. virginica* et *C. angulata*, seuls quelques gènes ribosomiaux ont été cartographiés.

1.4 Contexte de l'étude :

1.4.1 Les caractères d'intérêt pour l'amélioration chez les huîtres

Malgré l'importance économique de la production de l'huître creuse *Crassostrea gigas* en France et dans le monde, cette espèce a peu bénéficié des améliorations observées chez la quasi-totalité des espèces d'intérêt agricole et un nombre croissant d'espèces d'intérêt aquacole. Le principal apport de la génétique à l'amélioration de la production de cette espèce consiste en la production d'individus triploïdes (pour revue voir Nell, 2002). Chez les huîtres, les principaux objectifs des programmes de sélection ont été classés par Mahon (1983) en quatre groupes principaux par caractères d'intérêts : (a) caractères liés à la ponte et au développement larvaire ; (b) amélioration de la survie et de la croissance de la fixation à la taille commerciale ; (c) résistance à des facteurs environnementaux défavorables ; et (d) morphométrie de l'huître. La survie et la croissance sont toujours les deux caractères prioritaires étudiés dans les programmes de sélection (Lymbery, 2000).

Les huîtres sont, *a priori*, de très bons candidats pour des programmes de sélection compte tenu de leur importance économique, du contrôle de leur cycle biologique et de la grande variabilité génétique. En Australie, plusieurs programmes de sélection ont débuté dans les années 1990 visant à améliorer la croissance chez *Saccostrea commercialis*. A partir de 1990, un programme de sélection massale a été initié par le NSW Fisheries. Les résultats acquis après deux générations de sélection étaient encourageants (Net et al. 1996; Nell et al., 1999). Pour *C. gigas*, un programme de sélection a débuté dans ce pays en 1996. Les principaux objectifs de ce programme visent à améliorer la croissance, la qualité de la chair et la forme de la coquille par l'intermédiaire de plusieurs stratégies de sélection (sélection individuelle et/ou familiale, croisements consanguins). Les premiers résultats indiquent un gain de performance pour tous les caractères étudiés avec les meilleurs résultats via la sélection familiale (CSIRO, 2002). Aux USA, le « Molluscan Breeding Program » (<http://www.hmsc.orst.edu/projects/mbp>), programme de sélection ayant pour objectif d'améliorer le rendement en culture, a été initié en collaboration avec les écloséries américaines sur la côte ouest. Les résultats de ce programme, basé sur un nombre important de familles (>400), montrent un gain génétique significatif pour ce caractère (Langdon et al., 2003).

L'héritabilité désigne la part des gènes dans la variance d'un caractère exprimé. Il s'agit d'une mesure statistique et quantitative qui ne concerne jamais un individu, mais toujours un

ensemble d'individus. L'estimation de l'héritabilité et des corrélations génétiques entre caractères d'intérêt apparaît un préalable indispensable au lancement d'un programme de sélection de ces caractères. Chez les bivalves, un nombre relativement important d'estimations d'héritabilité pour divers caractères est rapporté dans la littérature, mais la grande majorité de ces études concerne la croissance et très peu la survie. En effet, ce caractère est difficile à améliorer par l'intermédiaire de la sélection génétique (Gjedem, 1985), alors que la croissance est généralement considérée comme étant un caractère présentant une héritabilité faible à modérée (Toro et al., 1990).

1.4.2. Le caractère de survie aux mortalités estivales

Aux USA, les travaux menés chez *C. gigas* dans les années 1970-1980, suite aux problèmes de mortalités estivales survenues dès les années 1960, ont montré que ces mortalités affectaient différenciellement des familles bi-parentales (Beattie et al., 1980). Un schéma de sélection, basé sur (1) des « challenges » thermiques appliqués aux géniteurs et (2) la sélection des meilleures familles issues de ces géniteurs fut alors proposé (Beattie et al., 1980). Les résultats obtenus en troisième génération ont montré une diminution significative des mortalités (Hershberger et al., 1984) mais le schéma de croisement et de sélection utilisé n'a apparemment pas permis de calculer l'héritabilité du caractère ni de poursuivre à plus long terme ce programme. De la même manière, des différences significatives en terme de survie entre familles bi-parentales ont été mises en évidence en France (Pajot et al., 1998; Le Borgne, com. pers.). Cependant, les différences observées entre familles ne sont pas forcément d'origine génétique du fait (1) des biais possibles liés à la confusion entre effet « famille » et effet « environnement précoce » et (2) à la possible transmission verticale ou pseudo-verticale d'agents pathogènes (notamment le virus de type herpès).

Des travaux plus récents (Ernande, 2001) ont montré des résultats intéressants, notamment sur la survie d'huîtres âgées d'un an. Les bases génétiques de la croissance, de la survie et de l'effort reproducteur, ainsi que leur plasticité (capacité d'un génotype à exprimer différents phénotypes en fonction de l'environnement), ont été étudiées en se basant sur un plan de croisements de type hiérarchique (15 familles de plein-frères constituant 5 familles de demi-frères ; Ernande et al., 2004). Les résultats obtenus en milieu contrôlé et en milieu naturel ont montré des différences significatives entre familles pour l'ensemble des trois caractères étudiés. Parmi ceux-ci, la survie présente la plus forte héritabilité ($0,28^1 \pm 0,18$).

¹ L'héritabilité est comprise entre 0 (caractère non héritable) et 1 (caractère fortement héritable).

Pour ce qui concerne la plasticité de ces caractères, l'effort reproducteur présente une forte héritabilité ($0,58 \pm 0,33$). L'étude des corrélations génétiques et phénotypiques (en milieu contrôlé) entre les caractères de croissance, effort reproductif et survie est particulièrement intéressante. Les corrélations phénotypiques entre (1) croissance et survie et (2) croissance et effort reproductif sont négatives. A l'inverse, la corrélation phénotypique entre croissance et effort reproductif est positive. Le signe des corrélations génétiques entre ces caractères change en fonction des environnements dans lesquels les huîtres sont élevées. De plus, une forte corrélation génétique positive entre survie et plasticité de l'effort reproducteur a également été montrée. Ces résultats démontrent la grande flexibilité des « mécanismes » physiologiques sous-jacents et l'importance des interactions animal-environnement dans les études génétiques chez les huîtres.

1.4.3. Etude des bases génétiques de la mortalité estivale du naissain et des possibilités d'amélioration de la survie par sélection dans le cadre du programme Morest

Suite à ces travaux, la partie génétique du projet Morest (<http://www.ifremer.fr/morest-gigas/>), visait à déterminer si la survie du naissain de *C. gigas* en période estivale était héritable et donc sélectionnable. Les principales questions traitées étaient les suivantes :

- Existe-t-il une base génétique de la survie du naissain en période estivale ?
- Quelle est l'importance de cette base génétique si elle existe ?
- Quelle est la réponse à la sélection lorsque la survie est sélectionnée ?
- Quelle est la nature des corrélations entre la survie du naissain et d'autres caractères d'intérêt (survie ultérieure et croissance) ?

Afin de répondre à ces questions, un protocole expérimental de génétique quantitative a été mis en œuvre sur plusieurs générations afin de répondre à ces questions. Ces travaux ont également fourni également les bases nécessaires à la définition d'une stratégie de sélection génétique pour améliorer la survie et à sa mise en œuvre (Boudry et al., 2008)

1.4.3.1 Sélection en première génération (G_0 - G_1)

Un échantillon de 300 huîtres, issues de captage naturel de Seudre, a été prélevé en novembre 2000 dans le bassin de Marennes-Oléron. Ces huîtres ont été placées en maturation selon les conditions habituelles d'écloserie et reproduites aux cours de 3 séries successives de croisements. Pour chaque série, 6 mâles ont été utilisés et chaque mâle a été croisé avec

quatre femelles afin d'obtenir 6 familles de demi-frères et 24 familles de plein-frères (les mâles et les femelles étant choisis aléatoirement dans la population G_0).

La caractérisation sur estran des familles a été réalisée chez 17 familles de demi-frères (6 mâles x 3 séries mais perte d'une famille par mortalité en phase larvaire), dont chacune d'elle est constituée de 2 à 3 familles de plein-frères, dans un dispositif multi-sites composé de 3 sites ateliers : Ronce Perquis (« Ronce » dans le Bassin de Marennes-Oléron), Rivière d'Auray (« RA » en Bretagne sud) et Baie des Veys (« BDV » en Normandie). Ronce et RA sont des sites présentant habituellement les plus fortes mortalités de juvéniles entre juin et septembre, alors que BDV présente des mortalités estivales moins importantes à ce stade (Fleury et al., 2001).

Les mortalités entre la mise sur site et octobre ont été de 6, 16 et 25% sur les sites BDV, Ronce et RA respectivement. Les familles de demi-frères ont montré des mortalités moyennes de 2 à 34% en BDV, de 3 à 69% à Ronce et de 8 à 82% à RA (Figure 9). Tous les facteurs testés sont significatifs sauf les interactions site x mâle et site x femelle. La famille de demi-frères issue du mâle 14 a montré des fortes mortalités dans les 3 sites. Les familles de demi-frères issues des mâles 4, 7 et 8 ont montré de fortes mortalités (38 à 63%) à Ronce et RA. Inversement, les familles de demi-frères issues des mâles 2, 9, 10, 11, 15 et 18 ont montré des faibles pourcentages de mortalité (moins de 7% à Ronce, 12% à RA et 6% à BDV).

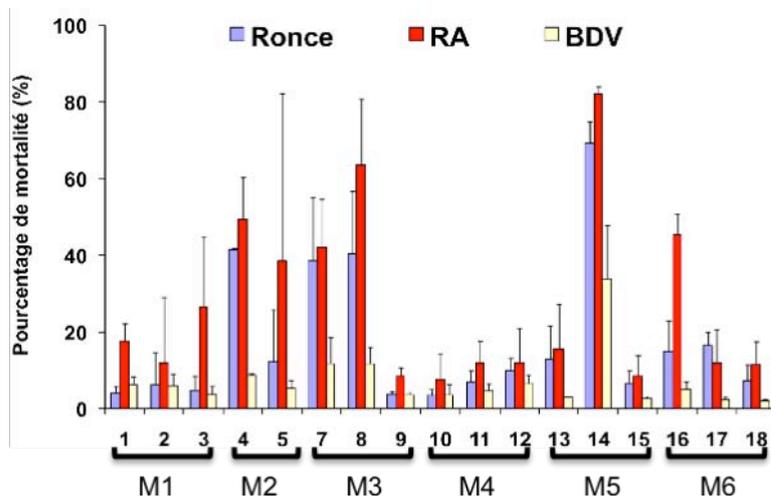


Figure 9 : Pourcentages moyens de mortalité et écart-types chez les 17 familles de demi-frères relevés dans les 3 sites, le 02 octobre 2001.

L'estimation d'héritabilité de la survie est très élevée sur le site de Ronce ($1,08 \pm 0,46$) et le site de RA ($0,89 \pm 0,40$) et modérée en BDV ($0,47 \pm 0,20$).

1.4.3.2 Seconde génération : sélection divergente pour une faible ou une forte survie du naissain sur le terrain en première période estivale (G_1 - G_2)

Suite aux résultats obtenus en 2001, la meilleure et la plus mauvaise des familles de demi-frères pour chaque série ont été retenues pour une expérience de sélection divergente sur la survie au stade naissain. Les familles sélectionnées n'ayant pas ou peu subi les mortalités estivales de 2001 ont été dénommées familles « résistantes », ou 'R', et les familles sélectionnées comme ayant présentées des mortalités importantes en 2001 ont été appelées familles « sensibles » ou 'S'.

Pour chaque famille G_1 sélectionnée, 500 individus ont été conservés à la nurserie Ifremer de Bouin pour grossissement intensif afin d'obtenir des animaux aptes à se reproduire dès l'âge de 9 mois. Ces animaux, indemnes de toute mortalité, ont été mis en élevage sur estran dans le bassin de Marennes-Oléron de fin septembre à mi-janvier. Début janvier 2002, aucune mortalité n'ayant été constatée, 60 individus par famille ont été conditionnés pour reproduction afin de générer les familles « G_2R » et « G_2S ». Chaque famille a été obtenue par le croisement d'une moyenne de 25 mâles et 25 femelles de chaque famille G_1 sélectionnée afin de représenter au mieux leur valeur génétique. D'autre part, ces mêmes familles G_1 ont été reproduites en consanguinité (intra famille de plein-frères et de demi-frères) générant les familles « G_{2cR} » et « G_{2cS} »

Les lots G_2 ont été placés sur estran en période estivale et la mortalité a été suivie. Les mortalités moyennes ont été de 13,8%, 24,1% et 11,5% à Ronce, RA et BDV respectivement. De plus, l'interaction entre la survie des lots et les sites est significative. Pour ce qui concerne les lots G_{2c} , la mortalité moyenne sur le site Ronce a été de 19,6%. Les lots G_{2cR} ont montré une mortalité significativement inférieure (3,8 %) aux G_{2cS} (43,3%).

1.4.3.3 Troisième génération : répétabilité de la réponse à sélection divergente pour une faible ou une forte survie du naissain en première période estivale (G_2 - G_3)

Afin de tester la répétabilité de la réponse à la sélection et des estimations des paramètres génétiques observées en seconde génération, une troisième génération a été produite en 2003. De manière similaire aux générations précédentes, 500 huîtres par lot G_2 ont été gardées indemnes de toute mortalité et soixante huîtres par lot ont été conditionnées pour leur reproduction. Chaque croisement a été réalisé avec une moyenne de 23 mâles et 23 femelles par lot parental G_2 . Les lots G_3S et G_3R ont été obtenus par croisement des G_{2cR} et

G₂cS. Chaque croisement a été réalisé dans les deux sens possibles (mâles x femelles et femelles x mâles) avec une moyenne de 23 mâles et 23 femelles par lot parental G₂c.

Des croisements intra-lots (donc générant de la consanguinité) ont également été réalisés. Une moyenne de 12 mâles et 12 femelles a été utilisée pour chaque lot G₂R, G₂S, G₂cR et G₂cS, générant respectivement les lots consanguins G₃cR, G₃cS, G₃c²R et G₃c²S (c² indiquant 2 générations successives de consanguinité). Pour les lots G₃, les protocoles d'élevage larvaire, de micronurserie et de nurserie utilisés pour G₁ et G₂ ont été appliqués à l'identique. Le testage sur estran des lots G₃R et G₃S n'a été réalisé que sur le site RA.

Les mortalités moyennes des lots G₃, sur le site de RA, est de 27.1% et 72.5% respectivement pour les lots G₃R et G₃S. Sur le site de Ronce, la mortalité des lots G₃c²R (34,6%) a été significativement inférieure à celle des G₃c²S (73,1%)

L'examen plus détaillé des mortalités des G₃cR et G₃cS montre que les lots issus du mâle 15 (G₀), qui avaient montré des faibles mortalités dans les générations précédentes, présentent une mortalité élevée, en comparaison à celle des descendants des mâles 2 et 9. De la même manière, les lots issus du mâle 7 se distinguent des lots issus des mâles 4 et 14. Ainsi, les lots issus des mâles 7 et 15 ont montré une mortalité du même niveau que le témoin et ils ne semblent pas être respectivement aussi « sensible » ou « résistant » au phénomène des mortalités estivales qu'estimé en G₁ (figure 10). Les estimations d'héritabilité et leur significativité sont présentées dans le tableau 3.

Tableau 3: Estimations d'héritabilité réalisée estimée en troisième génération pour le caractère de survie du naissain (octobre 2003 ; n= nombre de familles ayant servi à ces estimations).

site		R		S		R + S	
		h ²	n	h ²	n	h ²	n
G3	RA	1,62±0,58*	12	0,74±0,09***	12	0,93± 0,07***	24
G3c + G3c2	Ronce	1,69± 0,91	23	0,43± 0,10***	23	0,70± 0,09***	46

Seuil significatif : * P>0,05 ; ** P>0,01 ; *** P>0,001

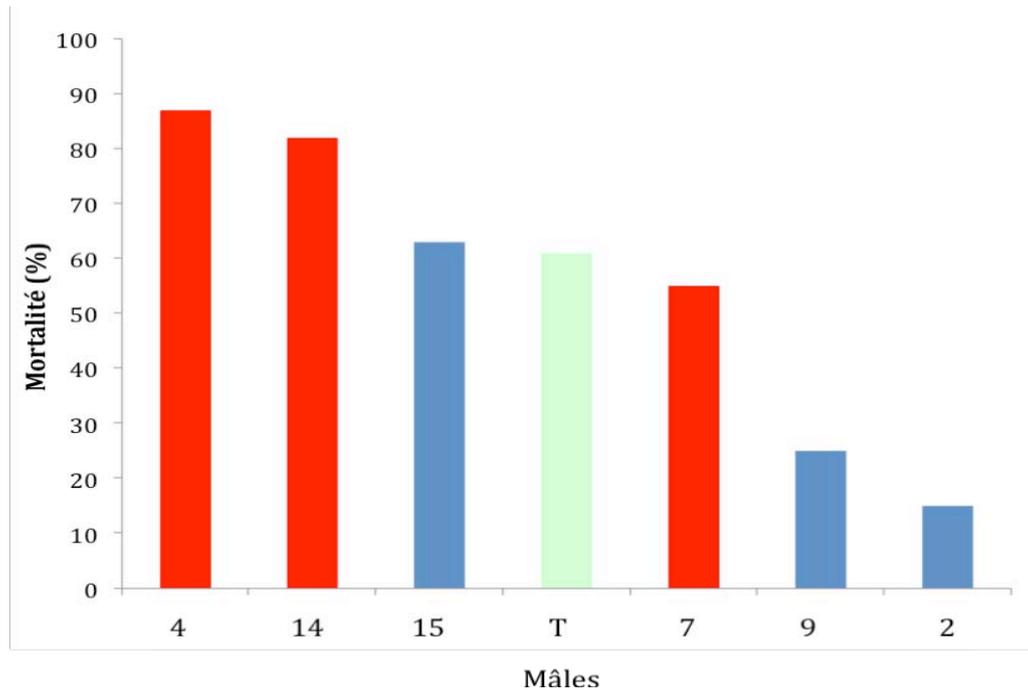


Figure 10 : Moyennes et écarts types par mâle de la génération «0» des mortalités des lots G₃cR et G₃cS et témoin (vert) en octobre 2003 sur le site de Ronce

1.4.3.4 Etude des performances de survie du naissain au laboratoire et corrélations avec les données obtenues sur estran.

En complément des suivis sur estran, plusieurs suivis ont été réalisés au laboratoire (bassins extérieurs et raceways intérieurs). Ceux-ci avaient initialement pour principal objectif d'identifier de manière précoce (c.a.d. avant les périodes de mortalité sur estran) des lots présentant des performances contrastées. L'ensemble des suivis réalisés est présenté dans le tableau 4. Pour les expériences L1 et L8, un stress thermique (2 h à 40°C en étuve) a été appliqué aux huîtres. Pour les expériences L3 et L4, contrairement à toutes les autres, le naissain a séjourné sur estran à Ronce entre la période de nursage à Bouin et les suivis en laboratoire. Les corrélations entre la survie au cours de ces suivis au laboratoire et les survies sur estran ont pu être examinées à partir des données présentées précédemment.

Tableau 4 : Ensemble des suivis réalisés pour l'étude de la survie en laboratoire

lots	Expérience	Croisement	Transfert en nurserie (Bouin)	Transfert sur estran	Début du suivi au laboratoire	Fin du suivi au laboratoire	Nombre de lots R testés	Nombre de lots S testés	Nombre de témoins 2n testés	Nombre de témoins 3n testés=
G2	L1	18/03	2/04	/	13/06	2/07	9	9	4	1
G2	L2	18/03	2/04	/	4/07	5/08	6	5	4	1
G2	L3	18/03	2/04	11/07	22/08	4/09	11	11	4	
G2c	L4	29/04	18/06	8/08	4/09	17/09	10	8	3	
G3	L5	17/02	10/04	/	25/06	4/09	12	12	1	3
G3c ²	L6	24/03	15/05	/	9/07	04/09	12	11	1	
G3c	L7	28/04	11/06	/	24/07	04/09	12	12	1	

- Résultats des suivis en laboratoire

Dans l'expérience L5, le naissain G₃ âgé de 4 mois a été testé en laboratoire à partir du 25 juin. Les mortalités ont eu lieu durant une semaine à partir du 23 août. Les mortalités des lots R ont varié de 19 à 42%, celles des lots S de 42 à 67% (témoin 2n : 52%). Les expériences L6 et L7 ont présenté des résultats similaires, les différences entre lots R et S étant toujours significatives.

- Comparaisons entre les mortalités sur estran et au laboratoire

Les mortalités des lots testés dans les expériences en laboratoire L1 à L8 ont été corrélées à celles observées sur estran. En 2003, pour l'expérience L3, les corrélations sont significatives pour le site RA ($r = 0,63$, $p < 0,01$) et le site de Ronce ($r = 0,47$, $p = 0,01$). De plus, toutes les corrélations entre mortalité au laboratoire et sur estran sont significatives.

- L5 / RA (G₃) : $r = 0,85$, $p < 0,0001$
- L6 / Ronce (G₃c²) : $r = 0,84$, $p < 0,0001$ (Figure 11)

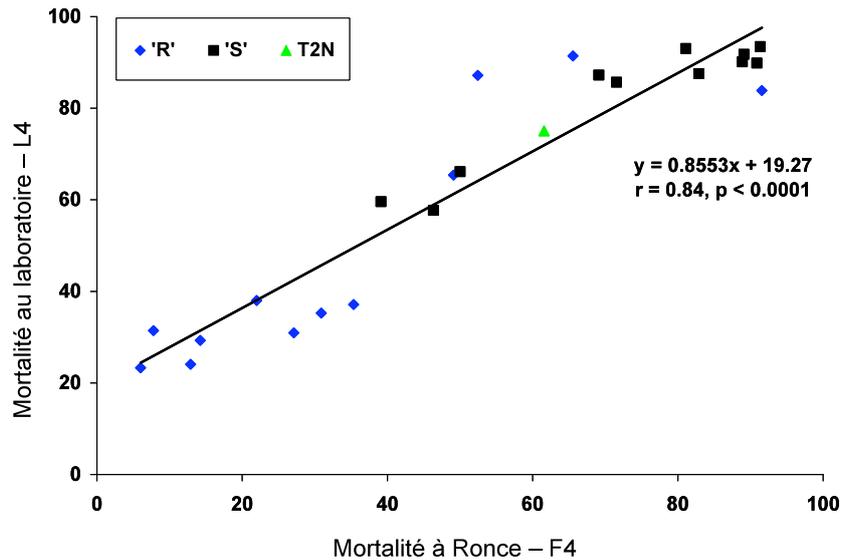


Figure 11 : Corrélation entre les mortalités cumulées au cours de l'expérience L4 et les mortalités à Ronce sur estran des lots G_3c^2 .

L'ensemble des résultats obtenus supporte l'hypothèse de l'existence d'une base génétique importante aux variations observées pour la résistance aux mortalités estivales : 46% de la variabilité totale observée dans la G_1 (Dégremont et al., 2005). Les résultats obtenus dans le cadre du programme de sélection « MBP » sont très semblables (Evans & Langdon, pers. com.). Les estimations de l'héritabilité du caractère obtenues en G_1 , les estimations d'héritabilité réalisées pour améliorer ou, au contraire, diminuer la survie estivale du naissain, sont très élevées. Ces résultats peuvent apparaître surprenants, les caractères fortement liés à la fitness présentant généralement une héritabilité faible ou nulle. La répétition en G_3 des résultats obtenus en G_2 pour la réponse à la sélection et la cohérence de nos résultats (entre sites, entre années, entre lots consanguins ou non consanguins et entre niveaux de ploïdie) est cependant à souligner. Globalement, nous pouvons observer que lorsque les mortalités sur les lots témoins sont supérieures à 10%, les lots R correspondants ont toujours une survie significativement supérieure (sauf un cas d'égalité pour les G_3c) et que les mortalités des lots S sont toujours significativement supérieures à celles des témoins (et également celles des R). Globalement, les différences observées entre les lots suivent les mêmes tendances qu'en premier été (quand les mortalités sont supérieures à 10%, les lots R présentent des survies significativement supérieures aux lots S, sauf pour la G_3 en RA).

Les résultats corroborent ceux précédemment observés aux USA (Beattie et al., 1980; (Hershberger et al., 1984) et en France (Ernande et al., 2004). La plus forte héritabilité du caractère dans ces expérimentations par rapport à ces travaux peut être liée au fait que ceux-ci

concernent des huîtres plus âgées (un an et plus). Si des mortalités ont lieu avant la période de mesure du caractère (ce qui est souvent le cas) et qu'il existe une base génétique à ces mortalités, la conséquence est une perte de la diversité génétique (par effet de la sélection) et donc des estimations d'héritabilité plus faibles. De plus, les conditions d'élevage n'étant jamais strictement homogènes entre les individus, plus le caractère est étudié tardivement, plus il peut être affecté par ces variations, sous-estimant ainsi sa base génétique.

La variabilité génétique, mise en évidence en G_1 et sélectionnée dans les générations suivantes, apparaît nettement additive. Ainsi, la réponse à la sélection pour augmenter la survie est très semblable à celle pour la diminuer. La consanguinité ne semble pas affecter de manière significative la survie du naissain aux mortalités estivales. Les suivis de lots consanguins et non consanguins effectués à Ronce en G_2 montrent des mortalités similaires. Les effets de la consanguinité ont été étudiés chez de nombreux bivalves : *Ostrea edulis* (Bierne et al., 1998), *Crassostrea virginica* (Longwell et al., 1973), *Pecten maximus* (Beaumont et al., 1983), *Argopecten circularis* (Ibarra et al., 1995), and *C. gigas* (Hedgecock et al., 1995) et les travaux publiés par Launey et Hedgecock (2001) ont clairement démontré un important fardeau génétique chez *C. gigas*.

L'expression du fardeau génétique en phase larvaire contribue à masquer l'effet de la consanguinité aux stades de développement ultérieurs (certains génotypes étant éliminés dès la phase larvaire). Nos résultats sont en concordance avec ceux publiés par Evans et al. (2004), qui montrent que la survie au stade adulte est relativement peu affectée par un niveau de consanguinité modéré. En G_3 les comparaisons sont moins directes car les lots consanguins ont été testés à Ronce et les non consanguins à RA, mais les lots G_3R et G_3S présentent des mortalités similaires aux lots G_3c^2R et G_3c^2S . Ces résultats pourraient être interprétés comme le masquage de la réponse à la sélection (observée dans les lots R et S non consanguins) par une « légère » dépression de consanguinité.

Parmi les différents facteurs permettant le maintien de variabilité génétique pour des caractères directement à la fitness, nous favorisons celle proposée par Ernande et al. (2004) : l'existence de « trade-offs » (effets de compensation) entre la survie et la reproduction. L'absence de réponse corrélative entre survie et croissance observée dans les expérimentations et les études de la reproduction entre lots R et S supportent cette hypothèse. L'investissement dans la reproduction est connu pour être très important chez l'huître adulte mais l'effort de reproduction chez le naissain a été moins étudié. Cette hypothèse ne sous-

entend pas forcément un lien physiologique direct entre coût de la reproduction et survie, l'intervention d'agents pathogènes comme agent causal des mortalités restant possible.

La résistance aux mortalités estivales n'en reste pas moins un caractère très plastique car très dépendant des conditions environnementales qui induisent (ou non) les mortalités. Cette importance de l'environnement sur l'expression du caractère peut contribuer au maintien de variabilité génétique du fait des variations, dans l'espace et dans le temps, des pressions de sélection (Ernande et al., 2004).

1.5 Objectifs généraux du travail de thèse

Forts des connaissances acquises, et du matériel biologique disponible, un projet européen (« Aquafirst : Combined genetic and functional genomic approaches for stress and disease resistance marker assisted selection in fish and shellfish » FP6, n°513692) a été initié. L'objectif global de ce projet est d'identifier les gènes associés au stress et la résistance à des maladies chez l'huître creuse, la truite, la dorade et le bar, afin de fournir une base physiologique et génétique pour la sélection assistée par marqueurs. Ces études ont été réalisées pour les poissons (bar, daurade, truite arc en ciel) en s'intéressant à des caractères de stress et de résistance à différents agents pathogènes directement liés à l'aquaculture, et chez l'huître creuse en utilisant des conditions qui induisent de fortes mortalités en période estivale.

Un des objectifs principaux de ce projet était d'identifier et caractériser des marqueurs liés à la résistance à la mortalité estivale chez *C. gigas*, en vue de leur possible application dans le cadre d'une sélection assistée par marqueurs.

Mon travail de thèse s'inscrit dans le cadre de ce projet. Il avait pour objectifs généraux de développer de nouveaux marqueurs moléculaires de type SNP, de les intégrer dans la construction d'une carte génétique consensus qui nous servira de base pour la recherche de QTL liés à la résistance à la mortalité estivale chez l'huître creuse *C. gigas*. Pour permettre une primo-localisation de ces QTL, nos efforts se sont concentrés sur quatre points :

- Produire le matériel biologique adapté (familles F2) et réaliser un suivi de mortalité en conditions expérimentales (phénotypage)

- Développer de nouveaux marqueurs moléculaires de type SNP et microsatellites à partir des séquences EST disponibles dans le domaine public

- Génotyper et intégrer des marqueurs précédemment cartographiés par Hubert et Hedgecock (2004) et les marqueurs nouvellement développés pour construire d'une cartographie génétique consensus

- Réaliser une primo-localisation de QTL(s) de résistance à la mortalité estivale

Ce document de thèse se présentera donc selon les quatre points indiqués précédemment. Ainsi, le chapitre II aborde le schéma de production de familles ségrégeantes F2. Ces familles ont été placées dans des conditions expérimentales permettant de suivre l'évolution temporelle des mortalités. Ce suivi nous a permis de récolter des données phénotypiques qui ont ensuite été combinées aux données de génotypes (chapitre V) pour rechercher des associations génotype-phénotype pour le caractère survie à la mortalité estivale et la charge virale en herpès virus associée aux mortalités.

Le troisième chapitre est consacré au **développement de nouveaux marqueurs** de type SNP, avec une revue des méthodes de détection et de génotypage à « haut débit » de ces marqueurs et le développement de microsatellites par une approche « *in silico* » ou bioinformatique. Le **génotypage de l'ensemble de ces marqueurs** y est également décrit.

Les méthodes et résultats de la **construction d'une carte génétique consensus**, c'est-à-dire résultant de l'analyse combinée des données des différentes familles F2, sont décrits dans le chapitre IV.

Enfin, le chapitre V traite de la **recherche des associations entre génotype et phénotype** (ou QTL), pour les caractères de survie et de charge virale associée à la mortalité estivale chez l'huître creuse *C. gigas* et de **leurs effets et localisation**.

CHAPITRE II

CARACTERISATION
PHENOTYPIQUE DE
FAMILLES EN PERIODE
ESTIVALE

Ce chapitre a pour but de décrire la production de familles en ségrégation et de caractériser (dynamique de l'événement de mortalité) une expérience de suivi de mortalité en période estivale en vue de recueillir des données de phénotypes sur ce matériel. Ce matériel biologique servira ensuite de base à la construction d'une carte génétique consensus puis l'ensemble de ces données sera combiné pour rechercher des QTL liés à la variance des caractères phénotypés au cours de l'expérience.

2.1 Production de familles en ségrégation

Afin de réaliser une recherche de QTL, la construction d'une carte génétique est nécessaire. Ceci est réalisé à partir d'un matériel biologique adapté, chez lequel les caractères d'intérêt et les marqueurs sont en ségrégation. Selon Vaiman (2000), une famille idéale pour construire une carte génétique est une famille de trois générations d'au moins 40 descendants. Alors que de telles familles sont difficiles voire impossibles à constituer chez beaucoup de mammifères, elles peuvent être aisément produites chez des espèces beaucoup plus fécondes, comme les plantes, les poissons et les mollusques. En théorie, pour couvrir un génome de mammifère, 300 méioses informatives suffisent pour construire une carte comprenant environ 200 marqueurs répartis uniformément sur la carte de liaison (Elsen et al., 1994). Cela revient à disposer de familles informatives de 150 individus. Les types de descendance (« pedigree ») sont relativement divers, certains n'étant applicables que chez les végétaux :

- une descendance de première génération (F1 ; Chistiakov et al., 2008)
- une descendance de seconde génération (F2), issue de l'autofécondation d'un hybride simple F1 (Tanksley et al., 1992) ou issue du croisement de deux parents F1 (approche classique)
- une population F3 dérivée d'individus F2 par une génération d'autofécondation (Beavis et Grant, 1991 ; Gradiner, 1993)
- une population de lignées recombinantes dérivée des individus F2 par plusieurs générations d'autofécondation sans sélection (Reiter et al., 1992; Philipp et al., 1994)
- une descendance de croisement en retour ("backcross") de première génération lorsque l'hybride F1 est recroisé avec un de ses parents (Gebhardt et al., 1991; Durham et al., 1992)
- des lignées issues d'haploïdes doublés chez des espèces où l'androgenèse ou la gynogenèse sont réalisables (Graner et al., 1991; Ferreira et al., 1994)
- une famille d'hybrides simples, inter ou intra-spécifique (Grattapaglia et al., 1994)
- une famille d'hybrides doubles (Jarrel et al., 1992; Devey et al., 1994)

Le choix des parents à croiser et la création d'une descendance en ségrégation constituent la première étape de la cartographie, support de la recherche de QTL. Une fois les parents déterminés, une descendance en ségrégation d'un nombre aussi important que possible d'individus est nécessaire pour obtenir un échantillon d'événements méiotiques permettant d'estimer le plus précisément possible les taux de recombinaison entre les marqueurs. De plus, si le génotype d'un ou plusieurs grands-parents est connu, cela permet de déterminer de façon certaine la phase de liaison des allèles. Dans ces conditions, les individus recombinants et non recombinants entre deux marqueurs liés peuvent être cumulés même s'ils proviennent de familles différentes. En effet, l'analyse combinée de données de génotype et de phénotype de deux à n familles, offre un potentiel pour extraire des informations supplémentaires de ces données. Par exemple, l'analyse conjointe de deux à n populations ou familles permet d'obtenir plus de puissance pour détecter les QTL, qui n'apparaissent dans aucune étude individuelle ou peut être utilisé pour confirmer la présence de QTL détectés dans une seule population (Lander et al., 1995). L'analyse conjointe permet des estimations plus précises des effets et de la localisation de QTL et pourrait être utilisée pour étudier les différences d'effets de QTL entre différentes populations (Walling et al., 2000).

De nombreuses cartes génétiques ont été réalisées à partir de croisements de plantes appartenant à des espèces différentes, ceci dans le but d'augmenter le nombre de marqueurs polymorphes. Des cartes génétiques ont ainsi pu être obtenues chez la tomate et la pomme de terre (Tanksley et al., 1992; Haanstra et al., 1999) ou le coton (Lacape et al., 2003). Ces espèces, bien que génétiquement différentes, produisent des hybrides F1 fertiles. Plus les espèces sont éloignées, moins leurs génomes seront susceptibles de recombiner, générant ainsi des cartes de liaison plus courtes qu'en croisement intra-spécifique. Le choix d'un type de descendance dépend essentiellement des objectifs de l'étude et des contraintes biologiques de l'espèce. Les familles issues de "backcross" ou les F2 sont les plus utilisées car elles sont relativement rapides à obtenir.

Il est admis que pour un même nombre d'individus analysés, une descendance F2 est préférable à une descendance issue d'un "backcross" car elle fournit deux fois plus d'événements méiotiques exploitables dans le cas de marqueurs co-dominants (Beavis et al., 1991). Les parents d'une descendance en ségrégation peuvent aussi être choisis pour maximiser les différences phénotypiques des caractères d'intérêt, permettant ainsi d'accroître la variance expliquée par le QTL (Lander et Botstein, 1989). Cependant, la cartographie de QTL est beaucoup plus difficile chez des populations de type « outbred » parce que tous les parents ne seront pas informatifs. Un parent informatif est un parent pour lequel les marqueurs et les QTL ségrégent. Une méthode pour augmenter la puissance de l'analyse est donc de choisir les parents sur la base des génotypes des marqueurs pour maximiser le déséquilibre de liaison dans leur descendance.

L'huître creuse est une espèce qui se prête bien au schéma de production de familles pour la recherche de QTL, car c'est une espèce à forte fécondité, les croisements sont bien maîtrisés et le temps de génération est relativement court (1 an). Dans notre étude, les individus F0 sont issus des lignées sélectionnées pour augmenter (lignées « R ») ou diminuer (lignées « S ») la survie du naissain obtenus en troisième génération du programme MOREST (Dégremont, 2003). Ces deux groupes de géniteurs présentaient donc des performances contrastées pour le caractère « résistance au phénomène de mortalité estivale » pour lequel une forte héritabilité ($h^2=0,7\pm 0,34$) a été mise en évidence (Dégremont, 2003). Ce matériel biologique est donc a priori bien adapté à un schéma expérimental de détection de QTL pour ce caractère. Par contre, cette sélection divergente ayant été strictement familiale (i.e. taux de survie), la possibilité d'obtenir des valeurs phénotypiques discriminantes restait à démontrer au niveau individuel.

2.2 Implication d'organismes pathogènes dans les phénomènes de mortalité estivale chez l'huître creuse

L'un des objectifs principaux en pathologie des mollusques est d'identifier et d'étudier les organismes pathogènes capables d'induire une maladie et des mortalités. Chez l'huître creuse, *Crassostrea gigas*, il est décrit en France des phénomènes de mortalité récurrents en période estivale depuis de nombreuses années. Des travaux en pathologie semblaient d'autant plus indispensables chez cette espèce qu'il n'était pas rapporté en Europe, jusqu'à ces dernières années, de maladies identifiées chez l'huître creuse, à l'exception d'une infection virale causée par un herpes virus. Ce dernier a été suspecté dès 1992 comme à l'origine d'épisodes de mortalité massive chez les larves et au stade naissain (Renault et al., 1995;

(Renault et al., 2001). Ainsi, dans le cadre du projet MOREST, un certain nombre de travaux ont été entrepris afin d'explorer l'implication d'organismes pathogènes dans le phénomène de mortalité estivale de l'huître creuse. Des travaux ont été plus particulièrement menés en bactériologie. A l'issue du projet, aucun organisme pathogène n'est apparu comme systématiquement associé aux mortalités estivales sur estran (Nicolas et al., 2007). Cependant, des bactéries du genre *Vibrio* (Lacoste et al., 2001) et l'herpès virus OsHV-1 (Ostreid Hepes virus type1) ont été décrits comme à l'origine de certaines des mortalités observées en période estivale. Les bactéries du genre *Vibrio* (*V. aestuarianus* et *V. splendidus*) sont détectées dans plus de 50% des cas de mortalités et le virus herpès est très fréquemment détecté chez les huîtres de moins de un an. En effet, la recherche du virus en PCR réalisée entre 1997 et 2006, lors de mortalités anormales observées en période estivale chez des lots de naissain de *C. gigas* a montré une corrélation positive entre la détection du virus OsHV1 et ces événements de mortalités. Parmi les lots présentant des mortalités, la détection du virus OsHV1 est supérieure à 40% alors que la détection de ce virus est inférieure à 2% chez des lots analysés hors période de mortalité (Garcia et al., com. pers ; Pépin, com. pers.). Des résultats comparables ont été rapportés aux Etats-Unis (Friedman et al., 2005).

Jusqu'à présent, aucune recherche d'agents pathogènes n'avait été effectuée sur la base d'un pas de temps journalier de prélèvement lors d'un épisode de mortalité estivale chez l'huître creuse. Ceci s'explique en particulier par la difficulté technique de réaliser de telles analyses pour des animaux élevés sur estran essentiellement pour des raisons d'accessibilité (zones de balancement des marées). Compte tenu des récentes avancées techniques en matière de quantification de l'ADN du virus OsHV1 (Pépin et al., 2008), il apparaissait particulièrement intéressant de combiner (1) le suivi de mortalité mis en place pour le phénotypage des familles F2 (décrit ci-dessous) et (2) l'évolution temporelle de la détection et de la quantification de l'ADN viral associée à la mortalité observée grâce aux nouveaux outils de virologie moléculaire disponibles (PCR quantitative). Ces deux phénotypes (mortalité et quantité d'ADN viral par individu interprétée comme la charge virale) ont été mesurés au cours d'un épisode de mortalité de familles F2 pour être finalement intégrés dans nos analyses de recherche de QTL.

2.3 Production du matériel biologique

Afin d'optimiser la détection de QTL, nous avons retenu le schéma de type F2. Les précédents travaux de construction de cartographie génétique chez *C.gigas* ont été menés sur une descendance obtenue par backcross (Li et Guo, 2004) ou par la production d'une F2 (Hubert et Hedgecock, 2004). La production de familles F1 biparentales a eu lieu en juillet 2004 à la station expérimentale Ifremer d'Argenton. Le naissain a été transféré en nursage à l'écloserie Ifremer de Bouin (figure 12).

L'ensemble des phases de croisement, d'élevage larvaire et de pré-grossissement des familles F2 a été réalisé au sein de l'écloserie expérimentale Ifremer de La Tremblade. Un grossissement dit « intensif » du naissain a été assuré à la station Ifremer de Bouin grâce à la production en masse de la micro-algue fourrage *Skeletonema costatum*. Cette phase de grossissement s'est déroulée du 12 avril au 31 juin 2006 (soit douze semaines) et avait pour but de favoriser la croissance du naissain afin que la taille des individus soit la plus grande possible. En effet, le protocole expérimental prévu visait à récupérer des fragments de tissus en dégradation sur les individus morts au cours de l'épisode de mortalité estivale, ce qui nécessite d'utiliser des animaux d'une taille convenable (au moins supérieur à 4 cm) afin de récupérer un fragment de tissu de taille suffisante pour les analyses génétiques.

2.3.1 Choix des géniteurs

En 2006, sur la base des ascendants des individus F1, le choix des individus de la F1 a été réalisé afin de croiser les individus F1 les moins apparentés les uns aux autres. 150 animaux par famille F1 ont été ramenés de la station expérimentale de Bouin. Un échantillon de 50 individus par famille a été placé dans des conditions environnementales optimales pour leur entrée en maturation le 22 novembre 2005. Les 100 individus restants ont été placés en raceways pour disposer d'animaux de réserve. Au début de la phase de maturation, la température et l'apport trophique ont été progressivement augmentés pour atteindre, d'une part une température de 22°C, et d'autre part un régime algal composé de *Isochrysis aff. galbana* et *Tetraselmis suecica*, assurant une croissance optimale. Le croisement a été réalisé le 22 février 2006.

Lignées résistantes

Lignées Sensibles

			Mâle G0	4 et 14		4 et 14	
			Famille G1	F4-15*F4-16 / (F14-54) ²	(F4-16) ² / (F14-54) ²	F4-15*F4-16 / F14-55*F14-54	(F4-16) ² / F14-55*F14-54
			G2C1	B2*M2	D2*M2	B2*P2	D2*P2
Mâle G0	Famille G1	G2C1	G3SD	BM2 (K)	DM2 (Q)	BP2 (R)	DP2 (AC)
2 et 9	(F2-5) ² / (F9-36) ²	A2*O2	A02 (Y)	Y1K (R)	Y1Q (Z)	R1Y (N)	AC1Y (C)
	(F2-8) ² / (F9-36) ²	G2*O2	G02 (Z)	Z1K (B)	Z1Q (J)	R1Z (E)	AC1Z (H)
2 et 9	(F2-8) ² / (F9-35) ²	G2*J2	GJ2 (H)	K1H (F)	Q1H (V)	H1R (AC)	H1AC (O)
	(F2-5) ² / (F9/35) ²	A2*J2	AJ2 (V)	K1V (L)	Q1V (T)	V1R (Q)	V1AC (X)

Figure 12 : Plan de croisement à l'origine des individus F0 dans le cadre du projet Aquafirst.

Prenons l'exemple du croisement F1 : J*Q qui implique des individus des familles J et Q (cadres jaunes). Un individu de la famille J (entre parenthèses), de génération F0 dans le cadre du programme Aquafirst, est issu du croisement entre la femelle Z1 (le sexe femelle est codé par une lettre et un chiffre) et le mâle Q. La femelle Z1, de la génération G3SD du programme MOREST, est elle-même issue du croisement, en génération G2, des individus des familles G2 et O2 (lire à l'horizontal vers la gauche), eux-mêmes issus d'animaux issus de la lignée sensible au phénomène de mortalité estivale (mâles G₀ n°2 et 9). A l'opposé, le mâle issu de la famille Q est issu de la lignée résistante au phénomène de mortalité car il est issu du croisement entre des individus des familles D2 et M2 (lire en vertical et vers le haut), eux-mêmes descendants d'individus de lignée résistante (Mâle G₀ n°4 et 14).

2.3.2 Plan de croisement

Pour chaque paire de familles F1, quatre croisements bi-parentaux (ou répliqués) faisant intervenir quatre couples de parents différents ont été réalisés, permettant de faire face à d'éventuels problèmes de mortalités au cours des phases précoces du cycle de vie. Pour chacun des cinq croisements inter-familles, quatre répliqués ont été réalisés. Ainsi, la famille F2 a été nommée B*X car les individus de la famille F1 B ont été croisés avec des individus de la famille F1 X. Quatre croisements (notés de 1 à 4), impliquant deux mâles et deux femelles de chacune des deux familles, ont été réalisés (Tableau 5).

Tableau 5 : Plan de croisements F2

Croisement Inter-Familles	Répliqués
R * O	Mâle R1 * Femelle O1 Mâle R2 * Femelle O2 Femelle R3 * Mâle O3 Femelle R4 * Mâle O4
Z * AC	Mâle Z1 * Femelle AC1 Mâle Z2 * Femelle AC2 Femelle Z3 * Mâle AC3 Femelle Z4 * Mâle AC4
B * X	Mâle B1 * Femelle X1 Mâle B2 * Femelle X2 Femelle B3 * Mâle X3 Femelle B4 * Mâle X4
J * Q	Mâle J1 * Femelle Q1 Mâle J2 * Femelle Q2 Femelle J3 * Mâle Q3 Femelle J4 * Mâle Q4
E * T	Mâle E1 * Femelle T1 Mâle E2 * Femelle T2 Femelle E3 * Mâle T3 Femelle E4 * Mâle T4

2.3.3 Obtention des gamètes et fécondation

Les gamètes ont été obtenus par scarification de la gonade (« stripping ») à l'aide d'un scalpel puis récupérés et rincés à l'eau de mer. Les spermatozoïdes et ovocytes ont été filtrés sur des tamis de 25 et 60 µm respectivement. Un comptage, par analyse d'image grâce au logiciel SAMBA™, a été réalisé pour déterminer la concentration des gamètes pour chaque individu scarifié. Ce comptage est réalisé sur cellules de Malassez pour les ovocytes et cellules de Thoma pour les spermatozoïdes. Un ratio de 100 spermatozoïdes par ovocyte a été utilisé pour la fécondation. Les fécondations ont eu lieu dans des béchers de 500 ml contenant 100 ml d'eau de mer filtrée, à température ambiante (18-20°C) afin d'éviter toute compétition

spermatique. Pour chacune des jarres d'élevage de 30L, 3 millions d'ovocytes ont été fécondés et mis en élevage. Après mélange, les gamètes ont été laissés au repos pendant une période de trois heures. Suite à cette période, l'observation au microscope d'ovocytes fécondés a attesté du bon déroulement de la phase de fécondation. Les larves néoformées ont alors été transférées en salle élevage larvaire. En parallèle, des prélèvements de tissus ont été réalisés : pour chaque parent F1, la glande digestive et la gonade ont été découpées et éliminées et le reste des tissus a été conservé à une température de -20°C en vue de leur génotypage.

2.3.4 Elevage larvaire

Après la fécondation, les zygotes ont été placés en jarre de 30L d'eau de mer, filtrée à $0,4\ \mu\text{m}$ et chauffée à $23^{\circ}\text{C}\pm 2^{\circ}\text{C}$. Deux jours après fécondation, l'estimation du taux d'éclosion a été réalisée. Ces estimations sont basées sur le comptage de trois réplicats par jarre d'élevage larvaire et qui doivent comporter un minimum de 50 larves. Pour chaque lot ou croisement, l'ensemble de la population larvaire a été conservée, afin d'éviter l'impact des pratiques d'élevage sur la diversité génétique des lots. En effet, des pratiques zootechniques, telles que le tamisage sélectif (élimination des larves de petites tailles) a un impact significatif sur la diversité génétique des lots produits (Taris, 2005). Ce genre de pratique, dans un cadre expérimental, est donc fortement déconseillé car elles introduisent un biais important dans la représentation des génotypes de la population, ce qui peut engendrer des phénomènes de distorsion de ségrégation problématiques dans la construction de cartes génétiques (voir chapitre III).

2.3.5 Nurserie

Les premières larves oeillées retenues sur une maille de $220\ \mu\text{m}$, c'est à dire prête à se fixer, ont été transférées en micronurserie. Ces larves ont été placées dans un tamis de $150\ \mu\text{m}$, dont les parois intérieures sont recouvertes d'une pellicule de paraffine. Au préalable, une couche fine de microbrisure de coquilles d'huîtres a été disposée sur le fond du tamis. Cette brisure constitue un substrat propice à la fixation des larves. Les tamis ont été disposés dans des raceways dont l'alimentation est assurée par un flux continu d'eau de mer filtrée, chauffée à 23°C enrichie en phytoplancton. Les mêmes conditions ont été appliquées à l'ensemble des cohortes de fixation placées en micro-nurserie. Enfin, l'effectif maximum de larves placées par tamis est de 50000 individus. Lorsque l'ensemble des individus d'une cohorte de fixation

est retenu sur une maille de 2 mm, il est alors possible de transférer cette cohorte en nurserie (figure 13). L'étape de grossissement a été assurée par l'équipe technique sur le site Ifremer de Bouin pour assurer une croissance rapide des animaux. Cinq cohortes (Z3*AC3 ; R2*O2 ; B2*X2 ; J1*Q1 et E2*T2) ont été envoyées à Bouin le 21 avril 2006 pour une période de deux mois de nursage intensif : les animaux ont été placés dans des tubes tamis en raceways où une eau de mer chauffée et enrichie en phytoplancton circulait. Les animaux des cohortes restantes ont été placés dans des conditions moins propices à un développement rapide (apport trophique monospécifique, eau de mer non chauffée). Dès lors, ces lots constituaient une réserve permettant de modifier le choix des familles F2.

2.3.6 Suivi de la mortalité

A la fin du mois de juin 2006, les premiers échantillons (500 individus pour chacune des familles F2) ont été congelés à -20°C. Puis, un échantillon de 1000 individus par lot, a été mis en poche d'une maille de 4 mm, puis placé en raceway d'un volume de 120m³. En parallèle, 500 individus de ces cinq familles ont été mis en poche. Ainsi, pour chaque famille F2, deux poches, contenant respectivement 1000 (poche n°1) et 500 individus (poche n°2) étaient disponibles. Pour ne pas biaiser l'estimation du nombre d'individus morts qui ont été échantillonnés dans la poche 1, le prélèvement d'individus vivants a été réalisé dans la poche 2. Ce schéma expérimental nous a permis de comparer directement, la quantification d'ADN viral par PCR quantitative associée aux deux états des animaux (morts/vivants) à une même date. Chaque poche ostréicole a été vidée quotidiennement de son contenu pour estimer la mortalité et récupérer les individus morts. Tous les individus morts pour lesquels un fragment de tissu était encore prélevable (c.a.d. pas trop décomposé), ont été immédiatement individualisés puis congelé à -20°C, dans l'attente de l'extraction de l'ADN et de la quantification de l'ADN du virus OsHV-1 (charge virale). En plus de ces individus morts, un échantillon de 30 individus vivants a été prélevé quotidiennement. A la fin du suivi de mortalité (15 octobre 2006), 150 individus survivants par lot ont été échantillonnés. Cet échantillonnage constitue un groupe extrême de la distribution qui a été utilisé dans notre stratégie de génotypage sélectif que nous avons appliqué dans la recherche de QTL liés à la résistance au phénomène de mortalité observé en période estivale (voir chapitre V). Le second groupe extrême était représenté par les premiers individus morts au cours du suivi chez chacune des familles F2. De plus, la quantification de l'ADN viral a également été réalisée chez ces animaux survivants.

2.3.7 Extraction d'ADN

Pour chaque individu mort puis congelé, un fragment de tissu, préférentiellement de la branchie ou du manteau a été soigneusement découpé et isolé de la masse viscérale de chacun des individus. Ce fragment a été pesé au dixième de milligramme près. L'ADN est alors extrait à partir de ce fragment tissulaire selon le protocole recommandé du Kit Qiagen (Qiamp tissue mini kit® ; Robledo et al., 2000). Les extractions ayant donné un rendement trop faible ($>50 \mu\text{g d'ADN} \cdot \mu\text{L}^{-1}$) ont été répétées. Si la quantité d'ADN extrait a été à nouveau trop faible, l'échantillon a été remplacé par un autre. L'extrait brut d'ADN a été dosé par spectrophométrie (Biospectrophotomètre, Eppendorf) et visualisé par migration sur gel d'agarose à 1,5%. Ensuite, les solutions mères d'ADN ont été diluées en eau à diverses concentrations adaptées au génotypage de marqueurs microsatellites ($100\text{ng} \cdot \mu\text{L}^{-1}$), SNP ($50\text{ng} \cdot \mu\text{L}^{-1}$) et pour la quantification de la charge virale ($2\text{ng} \cdot \mu\text{L}^{-1}$).

2.3.8 Protocole de quantification de l'ADN viral par PCR quantitative

Les charges virales des individus morts, des individus vivants (c'est-à-dire échantillonnés au même moment) et des individus survivants ont été estimées par quantification du nombre de copies d'ADN viral par milligramme de tissu frais. Cette quantification repose sur une technique de PCR quantitative préalablement publiée (Pépin et al., 2008). Pour chaque échantillon, le mélange réactionnel (ou Mix) de PCR est détaillé dans le tableau 6.

Tableau 6 : composition du Mix de PCR quantitative pour la quantification du virus OsHV1

	Volume par échantillon (μL)
Mélange Full-Velocity	12,5
Amorce sens C9 [$2\mu\text{M}$]	2,5
Amorce antisens C10 [$2\mu\text{M}$]	2,5
Eau	2,5
ADN [$2\text{ng}/\mu\text{L}$]	5
Total	25

Le Mix Full-Velocity (Stratagene) comprend une Taq dite « Hot-Start² » ainsi que les tampons et réactifs (MgCl₂) nécessaires à son fonctionnement, des dNTP ainsi qu'un fluorochrome, le Sybr Green I[®]. Sous forme libre, dans le milieu réactionnel, ce fluorochrome n'émet aucune fluorescence. Dès lors qu'il se trouve en présence d'ADN double brin, il s'intercale dans la structure double brin et émet alors une fluorescence. L'intensité de cette fluorescence est proportionnelle au nombre de copies de l'ADN cible et permet une quantification. Cette fluorescence est mesurée par un appareil de PCR adapté (MxPro 3.0, Stratagene). La spécificité de l'amplification est contrôlée de manière systématique en vérifiant la courbe de dissociation (Bustin, 2000). Le couple d'amorces C₉-C₁₀, qui génère un amplicon de 197 paires de bases, cible la région C du génome viral codant pour deux protéines de fonctions inconnues (Arzul et al., 2001).

Chaque échantillon a été analysé en triplicat afin de pouvoir éliminer les valeurs aberrantes. En cas d'erreur technique, l'analyse systématique des échantillons en triple permet d'écarter un réplicat sans répéter l'analyse dans sa totalité. La charge virale d'un échantillon, pour lequel un réplicat montre une valeur « aberrante », est calculée avec les deux valeurs restantes. Le profil thermique appliqué au cours de la réaction de PCR est décrit en figure 14.

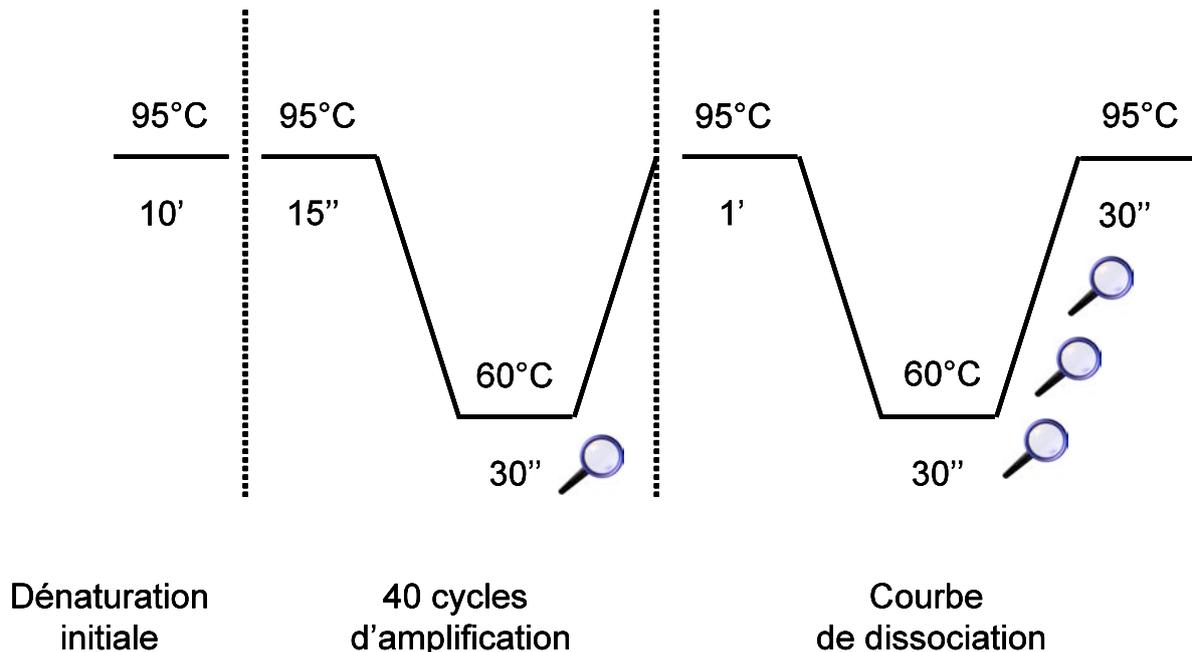


Figure 14 : Profil thermique appliqué au cours de la réaction de PCR-Q

² Taq polymérase modifiée qui s'active par une exposition de quelques minutes à une température élevée (94°C)

Chapitre II : Caractérisation phénotypique de familles en période estivale

Une gamme de dilution d'ADN viral (de 1.10^5 à 32 copies d'ADN viral/ μL) a été réalisée à partir d'une concentration d'ADN viral connue. Cette gamme de dilution permet d'établir une courbe standard servant de référence pour le calcul de la quantité d'ADN virale de chaque échantillon (Figure 15). Il est assumé qu'une particule virale contient une seule copie du génome du virus. Dans ces conditions, il est possible, à partir de la quantification de l'ADN viral, de définir la notion de charge virale (Arzul et al., 2001), c'est-à-dire le nombre de particules virales par unité de poids dans notre cas.

A l'issue des 40 cycles d'amplification, une courbe de dissociation est réalisée. Cette courbe permet de vérifier la spécificité du produit de PCR. La propriété d'intercalant du Sybr Green I[®] est exploitée. Au cours d'une montée en température lente, par paliers de $0,5^\circ\text{C}$, la fluorescence est mesurée toutes les secondes. Lorsque la température atteint la température de dissociation du produit de PCR, celui-ci passe d'une conformation double brin à une conformation mono brin. A ce moment, la fluorescence chute brutalement car le fluorochrome est libéré en solution.

La dérivé première de la courbe de dissociation révèle la température de dissociation du produit de PCR. Un pic représente la température à laquelle 50% des produits se sont dissociés. Cette courbe de dissociation révèle la spécificité de l'amplification. L'observation d'un second pic, à une température différente de celle du produit de PCR attendu, traduit la présence de structures en dimères ou de produits non spécifiques. Enfin, l'efficacité de la réaction de PCR et la linéarité (R^2), déduits de la courbe standard, sont vérifiées. En effet, une efficacité (E) trop faible ou trop forte ($90\% < E < 110\%$) ou une linéarité trop basse ($R^2 < 95$), biaisent les résultats de quantification. L'ensemble des paramètres analysés (E, R^2) est calculé par le logiciel dédié à l'appareil de PCR quantitative (Mx Pro 3.0).

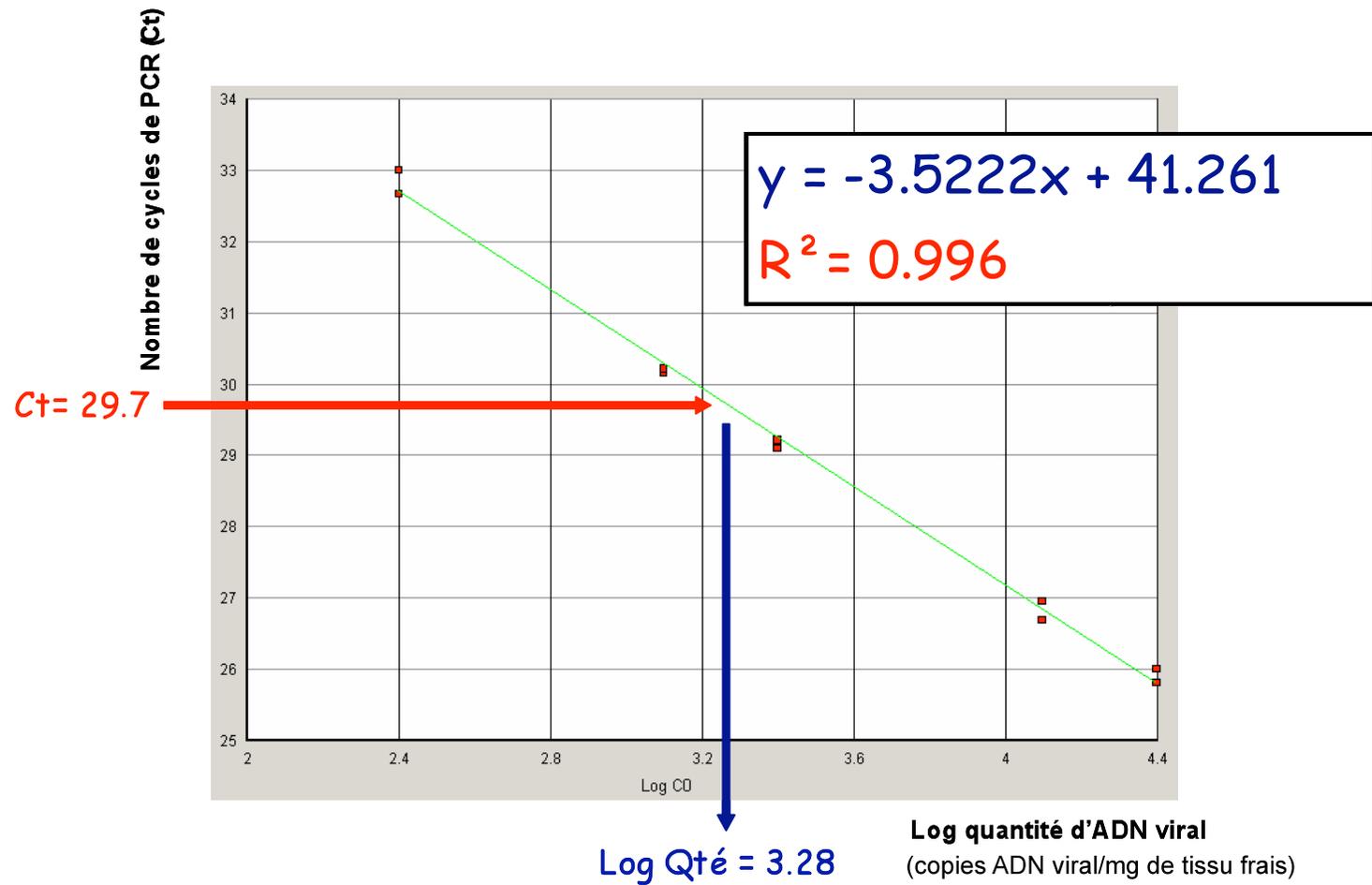


Figure 15 : Utilisation de la courbe standard pour déduire la quantité d'ADN viral présente dans l'échantillon

R^2 : linéarité

2.3.9 Analyses statistiques

Les statistiques descriptives (moyenne, maximum, minimum, médiane) pour les charges virales ont été obtenues par l'utilisation du logiciel de statistiques SPSS v11.5. La normalité des données a été examinée par le test de Shapiro-Wilk. De plus, la corrélation entre l'intensité de la mortalité estivale, estimée par le nombre ou le pourcentage d'individus morts par famille F2, et la charge virale associée a été mise en évidence, en utilisant le test du Rhô de Spearman. Le test de rang de Wilcoxon (test non paramétrique) a été utilisé pour mettre en évidence les différences de charge virale entre les individus morts et les individus vivants, échantillonnés au même moment, dans une même famille. Ce même test statistique a été utilisé pour mettre en évidence d'éventuelles différences significatives de charge virale moyenne des groupes d'individus vivants et morts entre familles.

2.4 Résultats

2.4.1 Elevage larvaire

Les résultats des taux d'éclosion à J+2 figurent dans le tableau 7. Mis à part le lot 9 (croisement B1*X1) qui a été perdu accidentellement (tamis déchiré), l'élevage larvaire s'est déroulé normalement et les premières larves oeillées sont apparues 22 jours après fécondation

Tableau 7 : Estimation du taux d'éclosion à J+2 et de pourcentage de larves fixées.

Lot	Familles	Effectif théorique	Effectif à J1	% Eclosion	Moyenne par famille	% de larves fixées
1	R1*O1	3000000	1232000	41,1	34,7	70
2	R2*O2	3000000	1447000	48,2		78
3	R3*O3	3000000	610000	20,3		67
4	R4*O4	3000000	878000	29,3		22
5	Z1*AC1	3000000	474000	15,8	27,3	64
6	Z2*AC2	3000000	48000	1,6		74
7	Z3*AC3	3000000	1212000	40,4		70
8	Z4*AC4	3000000	1536000	51,2		41
9	B1*X1	3000000	114000	3,8	27	- ^a
10	B2*X2	3000000	130000	4,3		100
11	B3*X3	3000000	1740000	58		42
12	B4*X4	3000000	1256000	41,9		20
13	J1*Q1	3000000	1236000	41,2	30,7	85
14	J2*Q2	3000000	46000	1,5		57
15	J3*Q3	3000000	1152000	38,4		62
16	J4*Q4	3000000	1246000	41,5		100
17	E1*T1	3000000	1384000	46,1	25,6	20
18	E2*T2	3000000	816000	27,2		81
19	E3*T3	3000000	554000	18,5		3
20	E4*T4	3000000	320000	10,7		100

-^a : pas de données disponibles

2.4.2 Taux de fixation et micronursage

Le taux de fixation moyen est de 60% avec un taux de fixation minimum et maximum de 3 et 100% pour les croisements E3*T3 et B2*X2 respectivement. L'ensemble des résultats des taux de fixation figure dans le tableau 7.

2.4.3 Phase de nursage

Lorsque le naissain a atteint une taille supérieure à 2 mm, l'envoi des cohortes de fixation, vers le site Ifremer de Bouin, s'est déroulé en deux temps :

1. Une cohorte (ou réplicat d'un croisement inter-famille) par famille F2 a été choisie selon deux critères : le succès à la métamorphose et la taille moyenne des individus.
2. Les individus restants ont également été envoyés dans leur ensemble vers la station de Bouin 64 jours après fécondation.

Des mortalités massives (>90%) sont survenues pendant la phase de nursage sur certains lots. L'une des cohortes (E2*T2), placée en nursage intensif, a été infectée par OsHV1, (détection du virus en PCR quantitative). La cohorte E2*T2 a été alors remplacée par la cohorte E4*T4 qui n'a pas subi de mortalité au cours de cette étape. Après une période de douze semaines de grossissement, le naissain des cohortes E₄*T₄, R₄*O₄, B₂*X₂, J₁*Q₁ et Z₃*AC₃ a été ramené sur le site de La Tremblade 130 jours après fécondation pour être placé en condition expérimentale. Par la suite, les familles F2 seront nommées, dans un souci de simplicité de lecture, uniquement par des lettres (ex : Famille E₄*T₄ sera nommée E*T).

Chapitre II : Caractérisation phénotypique de familles en période estivale

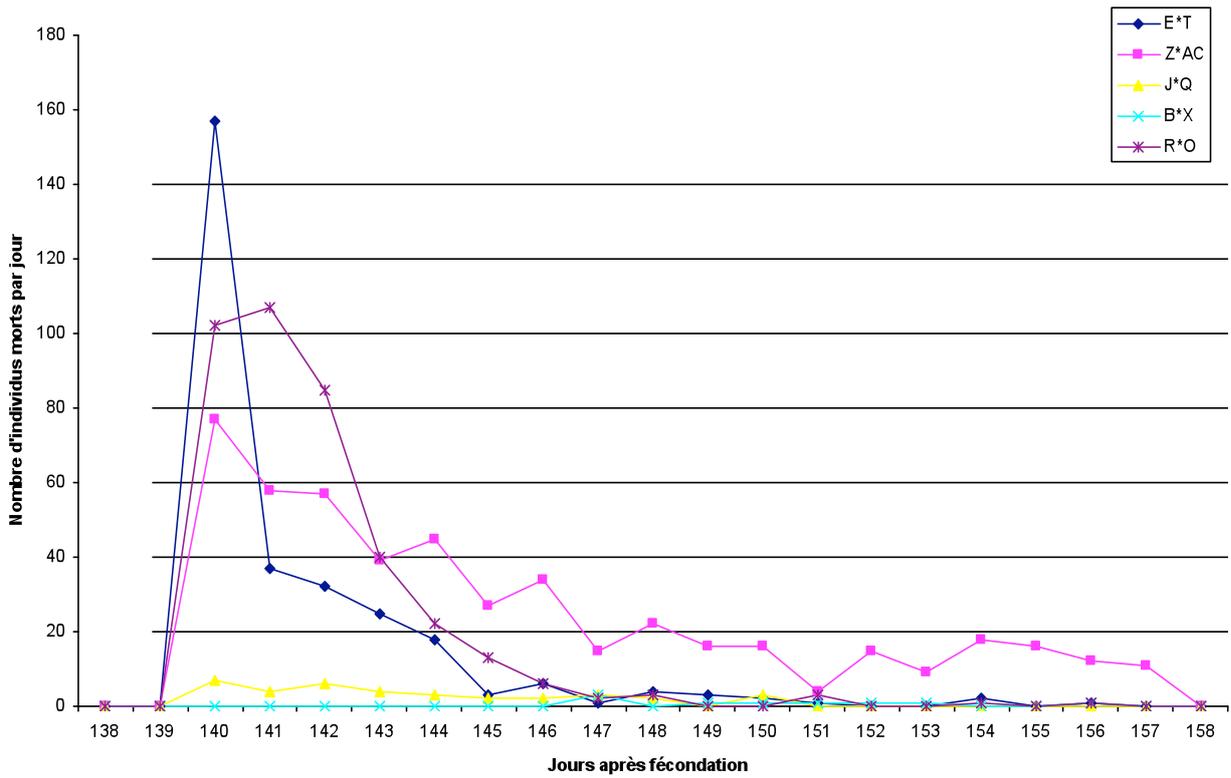


Figure 16 : Cinétiques de mortalité journalière des cinq familles F2 étudiées au cours de la période estivale 2006 (n=1000/famille).

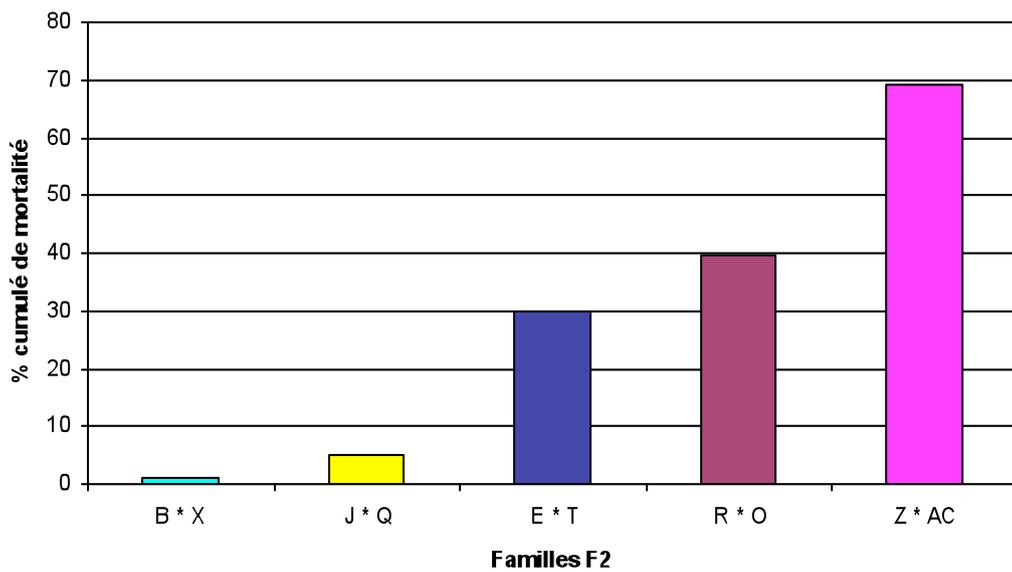


Figure 17 : Pourcentages cumulés de mortalité des familles F2 étudiées au cours de la période estivale 2006 (Juillet à octobre 2006, n=1000/famille).

2.4.4 Suivi de mortalité

Le suivi de mortalité a débuté 138 jours après fécondation et les premières mortalités massives sont apparues dès le 10 juillet. Au cours de la première semaine de suivi, 28, 35 et 37% des individus sont morts pour les familles E*T, Z*AC et R*O respectivement (Figure 16). A l'issue du suivi (15 octobre 2006), le pourcentage de mortalité cumulé maximum était de 67% (famille Z*AC). A l'opposé, les familles B*X et J*Q, placés dans le même raceway, n'ont subi que peu de mortalité : 1,2 et 5% respectivement (Figure 17). Pour la suite des analyses génétiques (génotypage de marqueurs moléculaires, cartographie de liaison et recherche de QTLs), seules les 3 familles E*T, Z*AC et R*O ont été étudiées. Le tableau 8 résume le nombre, la nature et la date des échantillons prélevés au cours de ce suivi de mortalité chez chacune des familles F2.

Tableau 8 : Echantillons prélevés puis congelés au cours du suivi de mortalité

Familles F2	Nbr d'individus morts au cours du suivi (%) (n=1000)	Nbr d'individus vivants prélevés quotidiennement (n=500)	Nbr d'individus survivants prélevés (octobre 06- n=1000)
B*X	12	30	150
J*Q	46	30	150
E*T	298	30	150
R*O	393	30	150
Z*AC	670	30	150

2.4.5 Extraction d'ADN

L'ensemble des individus morts, vivants ou survivants échantillonnés a été traité de la même manière. Conformément à notre stratégie de génotypage sélectif (30% de la distribution divisée en deux groupes extrêmes de 15%), les extractions d'ADN ont été menées sur les 150 premiers individus morts et 150 individus ayant survécu, chez chaque famille F2.

2.4.6 Mesure de la charge virale individuelle

Préalablement aux mesures de charges virales chez des animaux morts, des tests ont été entrepris. Tout d’abord, les échantillons dont la quantité d’ADN extraite était inférieure à 50µg/µL et de qualité insuffisante (DO 260/280>1.8) ont été écartés. Ensuite, les ADN ont été contrôlés visuellement par migration sur gel d’agarose à 1,5% et des PCR de test ont été effectuées avec des échantillons pris au hasard afin de vérifier l’amplification des ADN extraits.

Tableau 9 : statistiques descriptives de la charge virale par famille F2 et par condition (Individus morts, vivants ou survivants)

	N	Charge virale* individuelle minimale	Charge virale* individuelle maximale	Charge virale* moyenne	Ecart type
Famille E*T - Ind. Morts	150	347	8,22E+07	3,99E+06	9,78E+06
Famille R*O - Ind. Morts	150	5	2,95E+07	2,04E+06	3,86E+06
Famille Z*AC - Ind. Morts	150	0	4,84E+08	1,14E+07	4,66E+07
Famille R*O - Ind.Vivants (Ech. quotidien)	107	0	6,57E+07	1,08E+05	6,70E+06
Famille Z*AC - Ind.Vivants (Ech. quotidien)	150	0	1,10E+08	1,71E+05	1,11E+06
Famille B*X - Ind.Vivants (Ech. quotidien)	96	0	3,40E+03	2,50E+02	1,34E+03
Famille E*T - Ind. Survivants (Ech en octobre 06)	150	0	2,35E+03	1,02E+02	2,48E+02
Famille R*O - Ind. Survivants (Ech en octobre 06)	150	0	1,71E+03	3,34E+02	1,47E+02
Famille Z*AC - Ind. Survivants (Ech en octobre 06)	150	0	2,46E+02	1,86E+02	2,54E+02

* Charge virale : nombre de copies d’ADN viral par mg de tissu frais

L’hypothèse de distribution normale des données a été testée par le test de Shapiro-Wilk. Cette normalité n’étant pas respectée, un test de Kruskal-Wallis (équivalent non paramétrique de l’ANOVA) a permis de mettre en évidence une différence significative (P<0,001) pour l’analyse de variance de la charge virale entre individus morts, vivants et survivants. La quantification de l’ADN d’OsHV1 chez les 150 premiers individus morts au cours de l’expérience montre que la charge virale moyenne la plus élevée est détectée chez la famille Z*AC (Z=27,25, P>0,001) avec une charge virale moyenne estimée à 2,24.10⁷ particules virales/mg de tissu frais. Le nombre de copies d’ADN viral/mg de tissu frais est cependant relativement variable, car il est compris entre 0 et 4,84.10⁸. De même, chez les

Chapitre II : Caractérisation phénotypique de familles en période estivale

individus vivants, échantillonnés quotidiennement, la charge virale moyenne la plus élevée est observée chez les individus de la famille Z*AC ($Z=14,39$, $P>0,01$). L'ensemble des statistiques descriptives est résumé dans le tableau 9 et représenté en figure 18. Il est à noter que les individus vivants de la Famille E*T n'ont pas pu être analysés pour des raisons techniques. Chez la famille B*X, aucune analyse sur des animaux morts n'a été effectuée car celle-ci ne présentait que de très faibles mortalités (pas d'animaux morts). De plus, aucune analyse sur des animaux survivants (en fin d'expérience) n'a été effectuée car nous sommes intéressés à l'évolution dans le temps la détection du virus après une épisode infectieux avéré (c'est à dire pour les familles qui ont présentés de fortes mortalités associées à de très fortes charges virales). Enfin, chez les animaux survivants (en fin d'expérimentation), la charge virale est comprise entre 1.10^2 et $3,3.10^2$.

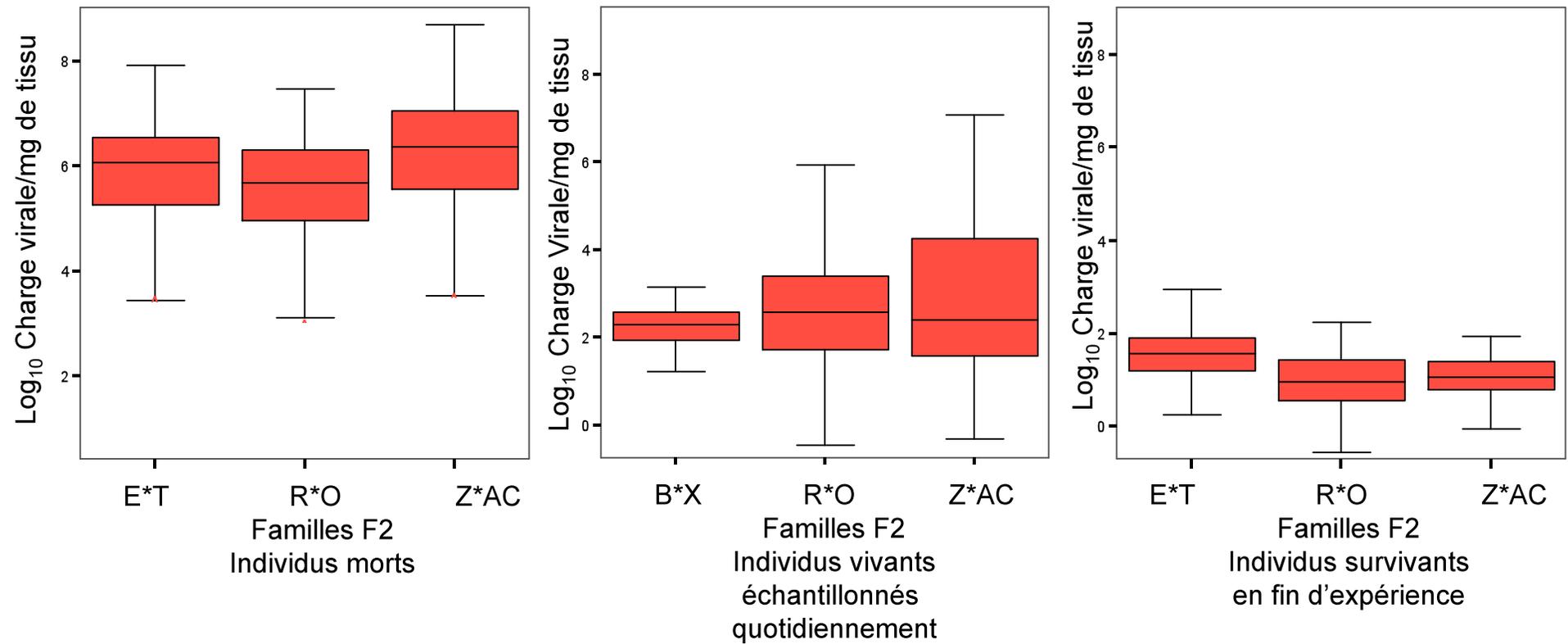


Figure 18 : Représentation graphique en Box-plot (Moyenne, écart-type, minimum et maximum) de la variance de la charge virale au sein des trois familles F2 et entre conditions

2.5 Discussion

Le suivi temporel serré, mis en place dans le schéma expérimental est le premier suivi qui permet de mettre en évidence l'évolution quotidienne de la charge virale moyenne chez les individus morts au sein d'un même lot et entre différents lots dans des conditions environnementales strictement identiques (même raceway). Les résultats obtenus démontrent une évolution très rapide de l'infection par l'herpès virus et permettent d'établir une nette corrélation entre les mortalités et la charge virale associée dans le cadre de ce suivi de mortalité. En effet, en l'espace de dix jours, l'infection virale s'est développée en affectant différenciellement les familles F2 et atteindre des charges virales individuelles supérieures à 1.10^8 copies d'ADN viral/mg de tissu frais. Cela confirme que la courte période dans laquelle les mortalités sont intervenues est compatible avec les mécanismes de réplication virale. De plus, nos résultats confirment la difficulté de suivre, de manière adéquate, l'évolution de ce virus chez des animaux placés sur l'étranger du fait de leur accessibilité limitée et de l'impossibilité de récupérer des échantillons de tissus sur ces animaux.

La détection différentielle du virus soulève l'hypothèse de son influence: (1) le virus OsHV1 est-il un marqueur d'un état de stress ; (2) le virus est-il l'agent causal de mortalité ? La première hypothèse paraît peu probable car nos résultats montrent une association forte entre le pourcentage de mortalité et l'augmentation de la charge virale.

La récupération quotidienne de tissus sur les individus morts et la mise au point de la technique de quantification de la charge virale par PCR quantitative, nous a permis d'établir plusieurs conclusions :

- (1) Parmi les individus morts au cours du suivi, la charge virale associée présentait une grande variance mais se corrélait avec l'intensité de la mortalité.
- (2) Les animaux vivants présentaient une charge virale significativement plus basse que les animaux morts. Cette observation supporte l'hypothèse de l'existence d'une base génétique sous jacente à la variation pour la survie estivale transmise entre générations.

Sachant que l'héritabilité du caractère résistance à la mortalité en période estivale est forte (Dégremont et al., 2007) mais que les lignées « R » et « S » utilisées en F0 ne sont pas fixées, une hypothèse possible pour expliquer les différences de survie moyenne observée entre famille serait que seules les familles R*O, Z*AC et E*T ségrégent pour le caractère, au contraire des familles B*X et J*Q. Ces différences seraient donc dues à des effets de tirage aléatoire des parents F0 utilisés dans les croisements. Alternativement, l'hypothèse d'un effet environnemental influençant différenciellement les familles (interaction génotype x environnement) est également possible mais moins probable.

Enfin, l'hypothèse du transfert vertical du virus OsHV-1 ne peut pas être exclue pour expliquer les résultats observés dans cette expérience. En effet, une transmission de ce virus des parents à leur descendance a été suspectée (Arzul et al., 2001; Barbosa-Solomieu et al., 2005; Boudry et al., 2008). Les différences observées entre familles en terme de mortalité pourraient être aussi expliquées par un portage différentiel du virus pour les géniteurs ayant donné naissance aux différentes familles. Les familles ayant présenté les très fortes mortalités pourraient être issues de géniteurs porteurs asymptomatiques du virus alors que celles n'ayant subi que de très faibles mortalités le seraient de géniteurs indemnes. Cependant, cette hypothèse à elle seule ne permet pas d'expliquer pourquoi les individus des familles B*X et J*Q, alors qu'ils ont été exposés au virus de la même manière que les animaux des autres familles n'ont pas contracté l'infection et exprimé la maladie. En effet, les individus des cinq familles ont été maintenus dans le même bassin tout au long de l'expérience et ont été exposés au même risque en terme de transmission horizontale du virus. Même s'il est reconnu que les virus enveloppés comme les herpès virus sont sensibles dans le milieu extérieur, il a été démontré pour tous les membres de la famille des *Herpesviridae* que la transmission a lieu d'animaux infectés à animaux sains. La présence d'un grand nombre d'individus morts dans le bassin au cours de l'épisode de mortalité et les charges virales très élevées détectées chez certains de ces individus laissent suspecter que toutes les conditions étaient réunies pour qu'une transmission horizontale survienne. De plus, de l'ADN viral a pu être détecté dans l'eau du bassin au cours de l'épisode de mortalité (Pépin, com. pers.) démontrant la présence du virus hors des animaux, même si la technique utilisée, la PCR quantitative, ne permet qu'une détection de l'ADN du virus et ne donne pas d'information sur son pouvoir infectieux. Dans ses conditions (transmission horizontale possible), les différences observées entre familles en termes de mortalité et de détection d'ADN viral s'expliqueraient par des capacités de résistance à l'infection virale différentielles. Il existerait une base génétique de la

résistance au virus OsHV-1 chez l'huître creuse, *Crassostrea gigas*. Cependant, il reste possible que les différences observées entre familles ne soient pas dues à une différence en termes de résistance à l'infection virale, mais bien à une capacité différentielle entre individus à transmettre l'agent infectieux (Renault, 2001). Une base génétique pourrait aussi exister pour ce phénomène. Le virus OsHV1 appartient à la famille des herpès virus qui se caractérisent, tout particulièrement par leur capacité de persister chez leur hôte sans se répliquer et induire de maladie (latence). Ces virus peuvent ainsi rester présents dans l'hôte toute sa vie durant. Ils n'induiront une maladie et des mortalités que lorsque que les conditions leur seront favorables (état de stress chez leur hôte). Cette stratégie est le résultat d'une longue co-évolution entre virus et hôte impliquant des interactions complexes. Dans ce contexte, il peut être très difficile de démêler l'écheveau concernant la résistance à la maladie entre capacités de défense différentielles et portage du virus.

La nature quantitative de la variabilité de la mortalité, mis en évidence par nos résultats, soutient l'hypothèse d'une base génétique. Les deux familles F2, n'ayant pas présenté de mortalité, ne sont pas informatives pour la recherche de QTL et ont donc été écartées des analyses suivantes (construction de carte génétique et recherche de QTL). Néanmoins, ces deux familles F2 sont intéressantes pour identifier une base génétique à la résistance au virus OsHV1. Le différentiel de mortalité observé entre familles pourrait apporter des éléments de compréhension des mécanismes de résistance à l'infection par le virus OsHV1. La variance phénotypique observée pour la charge virale associée aux mortalités, ainsi que la réponse différentielle des familles F2, vis-à-vis du caractère survie, laisse présager de la forte probabilité de détecter des régions chromosomiques (ou QTL) liées à la variance de ces deux caractères.

Au sein même de chacune des trois familles affectées par la mortalité, il existe une variance importante dans la charge virale détectée dans les échantillons de tissus. Les quantités d'ADN viral détectées chez certains animaux morts au cours du suivi, tendent à démontrer que le virus s'est répliqué de manière active chez ces individus ($>1.10^8$ copies d'ADN viral/mg de tissu frais). Ces derniers peuvent constituer des sources de virus, à partir desquels, ce dernier peut se transmettre aux individus voisins et les infecter à leur tour. De plus, la répllication active du virus laisse suspecter des lésions cellulaires et tissulaires extensives pouvant expliquer la mort des huîtres. Dans ces conditions, nous pouvons supposer que l'herpès virus est la cause majeure de mortalité dans le cadre de ce suivi. Néanmoins, la

recherche d'autres organismes pathogènes, tels que les bactéries du genre *Vibrio*, n'a pas été effectuée. De ce fait, l'implication agents infectieux autres que OsHV-1 dans l'épisode de mortalité ne peut pas être exclus. Des techniques moléculaires sont aujourd'hui disponibles au Laboratoire de Génétique et Pathologie (La Tremblade) pour explorer la présence d'autres organismes pathogènes associés aux mortalités, notamment les bactéries du genre *Vibrio*. L'analyse, avec ces outils, du matériel biologique phénotypé dans cette expérience, apportera un éclairage supplémentaire aux interactions et mécanismes qui interviennent pendant des évènements de mortalité.

2.6 Perspectives

D'un point de vue fonctionnel, l'étude de l'expression de gènes impliqués dans la réponse immunitaire chez l'huître creuse, à partir de techniques de biologie moléculaire, telles que l'hybridation suppressive soustractive (SSH), est une voie de recherche qu'il faut explorer. En effet, le matériel biologique, produit au cours de cette expérience, montre de forts contrastes en terme de réponse de la survie lors d'un épisode de mortalité survenant en période estivale. La mise en évidence de gènes préférentiellement ou spécifiquement exprimés ou réprimés, au cours d'une réponse déclenchée par une infection virale entre deux groupes de familles ou d'individus, est une piste prometteuse pour la compréhension des mécanismes de cette réponse. De plus, cela permettrait de (1) identifier des gènes clés dans ces mécanismes ; (2) établir des réseaux d'interactions entre ces gènes ; (3) d'intégrer *a posteriori* ces gènes comme marqueurs moléculaires dans une carte de liaison et vérifier la co-localisation de régions QTL avec ces marqueurs ou de mettre en évidence de nouvelles régions QTL.

Ces travaux ont débuté au laboratoire de Génétique et Pathologie (Ifremer, La Tremblade) par l'étude de l'expression, par PCR quantitative, de gènes ciblés. La fonction de ces gènes a été identifiée par homologie de séquence et montre que ceux-ci sont impliqués dans les mécanismes de réponse immunitaire. Des premiers résultats montrent une expression différentielle significative entre des animaux infectés et non infectés. A moyen terme, un plus grand nombre de gènes impliqués dans la réponse immunitaire liée à l'infection par l'herpès virus doivent être identifiés afin de préciser les connaissances concernant les mécanismes moléculaires et physiologiques sous jacent à cette réponse.

CHAPITRE III

DEVELOPPEMENT DE MARQUEURS MOLECULAIRES ET GENOTYPAGE

Ce chapitre aborde le développement de nouveaux marqueurs moléculaires SNP et microsatellites, et leur validation par génotypage chez les individus des trois familles F2. Ces marqueurs, ainsi qu'une sélection de marqueurs microsatellites, précédemment publiés, serviront de base à la construction d'une cartographie génétique (Chapitre IV).

3.1 Polymorphisme et marqueur moléculaire

La séquence nucléotidique d'un segment d'ADN, à un locus donné, dans le génome peut montrer certaines variations. Les différentes formes d'un locus sont appelées allèles. L'existence de différents allèles à un locus définit le polymorphisme génétique. Les techniques moléculaires qui révèlent le polymorphisme de l'ADN ont connu un développement très important ces dernières décennies. Avec l'augmentation du nombre des marqueurs, il est aujourd'hui possible de décrire ces variations génétiques à l'échelle du génome.

Un marqueur moléculaire peut présenter les caractéristiques suivantes :

- Spécificité : il peut être utilisé chez une ou plusieurs espèces, servant à les différencier et/ou à les comparer
- Mono- ou multi-locus (nombre de copies du marqueur dans le génome)
- Polymorphisme : un ou plusieurs allèles à chaque locus
- Réparti de manière aléatoire au sein du génome
- Neutralité ou non : soumis ou non aux effets de la sélection

Il est possible de distinguer trois catégories de marqueurs moléculaires en fonction du niveau d'information qu'ils apportent (Tableau 10):

- les marqueurs dominants : RAPD (Random Amplification of polymorphic DNA) et les AFLP (Amplified Fragment Length Polymorphism) ;
- les marqueurs co-dominants : ce sont les RFLP (Restriction Fragment Length Polymorphism) et SSCP (Single Stranded Conformation Polymorphism)
- les marqueurs multi-alléliques co-dominants tels que les VNTR (Variation Number of Tandem Repeat) qui sont les minisatellites, les microsatellites et SNP.

Tableau 10: Caractéristiques des principaux marqueurs moléculaires

Type de marqueur	Restriction enzymatique	PCR	Amorces Spécifiques	Gel	Caractéristiques techniques			
					Effort de développement	Effort de génotypage	Reproductibilité	Niveau d'information
RFLP	+	-	-	+	Important	Important	Important	Très important
PCR-RFLP	+	+	+	+	Important	Moyen	Important	Très important
RAPD	-	+	-	+	Très faible	Très faible	Faible	Très faible
AFLP	+	+	-	+	Faible	Très faible	Important	Moyen
SSCP	-	+	+	+	Moyenne	Moyen	Moyenne	Moyen
Microsatellite	-	+	+	+	Important	Faible	Important	Important
SNP	-	+	+	-	Important	Variable ⁽¹⁾	Important	Très important

(1) dépend de la technique employée (voir figure 20 – Lynch et Walsh, 1998)

Les marqueurs moléculaires de type dominants tels que les AFLPs sont, a priori, moins intéressants à utiliser de part leur nature et leur difficulté de transfert entre espèces et/ou entre laboratoires. Ils ont cependant l'avantage d'être relativement facile à développer en grand nombre. De plus, les AFLPs sont assez peu appropriés dans le cadre de la cartographie de QTLs (Quantitative Trait Loci) car ils ont tendance à se grouper. En effet, avec ce type de marqueurs, la détermination de la phase du marqueur est plus difficile. Cependant, ces marqueurs trouvent des applications telles que l'étude de diversité ou la construction d'une cartographie dense en marqueurs chez une espèce dépourvue de tous marqueurs (Van Haeringen et al., 2001). Depuis plusieurs décennies, les marqueurs moléculaires les plus couramment utilisés chez l'homme et les espèces animales agricoles telles que le porc (Rohrer et al., 1996) ou le poulet (Groenen et al., 2000) sont les microsatellites. Cependant, les SNP (« Single Nucleotide Polymorphism ») gagnent en intérêt et tendent à être utilisés à plus large échelle. Afin de mieux appréhender ce type de marqueur moléculaire, nous allons en décrire les principales caractéristiques, méthodes de détection et de génotypage ainsi que leurs principales applications. De plus, un bref comparatif entre marqueurs microsatellites et marqueurs SNP sera établi.

3.2 Les marqueurs microsatellites

Les microsatellites se définissent comme des séquences d'ADN composées de motifs répétés en tandem n fois. Ces motifs sont généralement composés de 2 à 6 nucléotides (ex : CG_n , $CAGC_n$, etc...) et peuvent être également appelés STR (Single Tandem Repeat) ou SSR (Single Sequence Repeat). Ces marqueurs sont communément décrits chez de nombreuses espèces eucaryotes et procaryotes, mais leur répartition au sein d'un génome n'est pas aléatoire (Toth et al., 2000). En effet, leur présence est rare dans les régions télomériques et centromériques des chromosomes (Wintero et al., 1992). De plus, l'étude de Toth et collaborateurs (2000) qui examine la répartition des microsatellites au sein de plusieurs taxons, montre que leur abondance est relativement importante avec un marqueur disposé tous les 2 à 10 kb selon le taxon considéré. Enfin, les microsatellites les plus communs sont les mono et dinucléotide ($n=1$ ou 2), composés principalement de bases A et T et leur taille n'excède pas 50 répétitions. Les marqueurs microsatellites présentent un fort polymorphisme dû à la variation du nombre de répétitions du motif qui les compose (Zhu et al., 2000). Ils présentent des propriétés de transmission mendélienne, de codominance, de relative neutralité et enfin de polymorphisme.

Leur développement peut s'opérer selon deux modalités : par séquençage de clones issus de banques enrichies (Zane et al., 2002). En général, cette modalité est appliquée chez des espèces dont le génome n'est pas séquencé. La seconde modalité consiste à la recherche bioinformatique de motifs dans des bases de données (méthode dite *In silico*) lorsque, chez l'espèce considérée, les ressources sont suffisantes et disponibles. Indépendamment de la modalité employée, lorsque la séquence de motifs est révélée par séquençage, un couple d'amorces est dessiné en vue de l'amplification d'un fragment contenant le marqueur au cours d'une réaction de PCR. Ce mode de développement est relativement long et coûteux. Cependant, cet effort est contre balancé par la facilité d'utilisation et la relative transférabilité (ou « cross amplification ») entre espèces très proches de ce type de marqueurs. Cependant, il existe une corrélation négative entre le succès d'amplification entre deux espèces proches et la distance génétique qui les séparent (Primmer et al., 1996).

De nombreux modèles de mutation ont été proposés afin d'expliquer l'important polymorphisme que présentent les marqueurs microsatellites. Les mécanismes de mutation se produisent essentiellement lors de crossing-over inégaux (Wolff et al., 1989) ou bien au cours de la réplication de l'ADN par phénomène de glissement de la polymérase suivi d'une absence de réparation (« Slipped Strand Mispairing » ou SSM, Levinson et al., 1987).

Le rôle fonctionnel de ce type de marqueurs n'est pas clairement défini. La plupart se situent dans les régions non codantes de l'ADN, raison pour laquelle aucune fonction ne pourrait leur être attribuée. Cependant, des études démontrent le rôle des microsatellites dans la régulation de gènes de part leur implantation dans la région promotrice de ces derniers (Li et al., 2002; Li et al., 2004) mais aussi l'organisation des chromosomes, la structure de l'ADN et l'organisation des centromères et télomères. Enfin, selon Kashi et collaborateurs (Kashi et al., 2006), les microsatellites ont un rôle important dans l'évolution des génomes, car ils permettent l'apparition et le maintien d'une variance génétique quantitative.

Les microsatellites sont, la plupart du temps, utilisés en tant que marqueurs moléculaires dans le cadre d'analyses de diversité de population, d'assignement de parenté (Figure 19) ou encore pour la construction de cartes génétiques. Cependant, leur utilisation requière une certaine attention car la possible présence d'allèles nuls, c'est-à-dire non amplifiés au cours de la PCR, peut induire des biais d'analyse. Cela est notamment vrai pour les loci hétérozygotes, qui en cas de non amplification d'un des deux allèles, sera observés comme homozygotes. Le biais engendré est alors relativement important dans la

caractérisation d'une population ou la recherche d'ascendants. L'explication la plus rationnelle à ce phénomène est l'apparition d'une mutation au niveau du site de liaison des amorces de PCR (Callen et al., 1993).

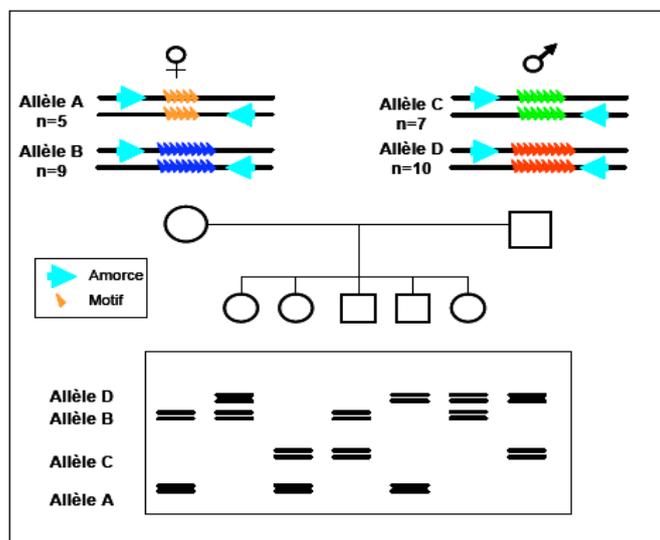


Figure 19 : Schématisation du principe de détection du polymorphisme de microsatellites en ségrégation dans des familles.

Chez l'huître creuse, le nombre de marqueurs moléculaires de type microsatellite est en constante augmentation grâce, notamment, au nombre croissant de séquences EST disponibles.

3.3 Revue des méthodes de génotypage des marqueurs moléculaires SNP

3.3.1 Caractéristiques des marqueurs SNP

Les SNP constituent la forme de variation la plus fréquente au sein du génome humain (Sachidanandam et al., 2001). Ce sont des marqueurs de choix de part leur abondance et leur stabilité supérieure aux microsatellites. Les SNP ont été principalement développés et utilisés pour cartographier des maladies complexes (Collins et al., 1998). Depuis quelques années, des SNP sont développés chez de nombreuses espèces d'intérêt comme le maïs (Thornsberry et al., 2001), l'arabette des dames (Cho et al., 1999), la canne à sucre (Grivet et al., 2003) ou le poulet (Ka-Shu Wong et al., 2004). A l'heure actuelle, plus de 1,4 millions de SNP ont été identifiés au sein du génome humain sous l'impulsion du programme international HGP (Human Genome Project). Les SNP sont une forme de polymorphisme engendrée par une mutation ponctuelle qui crée différents allèles à un même locus. Ces mutations ponctuelles, entraînant une différence dans une séquence d'ADN, ont été caractérisées depuis le début du

séquençage de l'ADN en 1977. Chez les mammifères, la fréquence de mutation est relativement faible puisqu'elle est estimée à 1×10^{-9} par site et par an à une position neutre (Li et al., 1981). La probabilité de modification d'une base à plusieurs reprises au même locus s'avère donc très faible d'autant plus qu'il existe un biais mutationnel. En effet, deux types de modifications existent : transition et transversion. Généralement, les transitions sont plus fréquentes que les transversions. La transition consiste en un passage d'une base purine à une autre base purine ($A \leftrightarrow G$) ou du passage d'une base pyrimidine à une autre pyrimidine ($C \leftrightarrow T$). La transversion consiste en un passage d'une base pyrimidine à une base purine ($A \leftrightarrow C$, $A \leftrightarrow T$, $T \leftrightarrow G$, $C \leftrightarrow G$). En théorie, un SNP peut présenter jusqu'à quatre allèles (les quatre bases possibles composant l'ADN) au même locus. Mais, la plupart des SNP sont considérés comme des marqueurs bi-alléliques car ils sont restreints à une ou deux bases. De plus, la fréquence de l'allèle le moins fréquent doit être supérieur à 1% pour qu'il soit considéré comme significatif. En tant que marqueur moléculaire co-dominant bi-allélique, le niveau d'information apporté par les SNP est inférieur à celui des microsatellites qui sont multi-alléliques. L'intérêt majeur des SNP réside dans leur abondance au sein de n'importe quel génome.

En plus de cette distinction, les mutations peuvent être qualifiées de synonymes, de non synonymes ou de non sens. Une mutation synonyme est une mutation qui n'entraîne pas de changement d'acide aminé à l'inverse des mutations non synonymes. Les mutations peuvent également introduire, de manière anticipée, un codon stop. Elle est alors appelée mutation non sens. Celle-ci est beaucoup plus rarement observée que les deux autres types de mutations car, à l'issue de la traduction, la protéine néo-synthétisée est généralement non fonctionnelle. Il y a un coût en terme de fitness élevé pour l'organisme lorsqu'un codon stop est introduit par erreur lors d'un événement de mutation. Enfin, une mutation est dite silencieuse si celle-ci est une mutation synonyme qui apparaît dans la partie codante de l'ADN mais aussi tout type de mutation dans les parties non codantes de l'ADN. Une mutation silencieuse n'induit donc aucun changement d'acide aminé.

Si les mutations SNP apparaissent de manière totalement aléatoire dans les génomes, le ratio transition/transversion devrait être égal à 0,5. Or, ce n'est pas le cas. La plupart des données de ce type montre clairement un biais mutationnel en faveur des transitions (Vignal et al., 2002). Par exemple, une étude a montré chez l'homme que le ratio transition/transversion était égal à 1,7 (Picoult-Newberg et al., 1999). En effet, la structure dégénérée du code génétique induit ce biais car la plupart de la diversité des codons est

composée de mutations synonymes. De ce fait, le ratio transition/transversion tend à augmenter dans les parties codantes de l'ADN. Chez l'huître creuse, ce ratio ($ts/tv=1.3$) est relativement proche de ceux observés chez d'autres organismes tels que la Drosophile (1.5 – Moriyama et al., 1996) ou l'homme (1,4 – Brookes, 1999). De plus, ce biais mutationnel s'applique à une échelle large car il existe généralement une forte corrélation entre la diversité nucléotidique non-synonyme dans le codant et le non codant (Moriyama et al., 1996).

Les SNP sont aujourd'hui des marqueurs largement utilisés de part leur abondance au sein des génomes. Leur utilisation relativement routinière, et la diversité des techniques de génotypage constituent les avantages principaux de ce type de marqueurs.

Les SNP sont retrouvés en moyenne tous les 1 kb chez la plupart des espèces (Kwok et al., 1996) telles que l'homme (Sachidanandam et al., 2001). Cet intervalle peut descendre à 1 SNP toutes les 125 paires de bases (pb) chez le moustique *Aedes aegypti* (Morlais et al., 2003), 1 SNP toutes les 104 pb chez le maïs *Zea mays ssp. mays L.* (Tenaillon et al., 2001) et un SNP toutes les 40 pb au niveau des régions non codantes du génome de l'huître creuse (Curole et Hedgecock, 2005). Mais, le niveau le plus élevé de polymorphisme reporté à ce jour a été détecté chez le ver rond *Caenorhabditis remanei* avec un SNP toutes les 20 paires de bases sur l'ensemble de son génome (Cutter et al., 2006).

Il existe un autre type de marqueur nucléotidique. Ce sont les insertions/délétions (« indels ») qui génèrent une variation de longueur de la séquence nucléotidique (figure 20). Les « indels » sont plus fréquents au niveau des introns que des exons, car ces derniers sont des régions relativement bien conservées. Les « indels » de grande taille sont plus facilement identifiables par la comparaison de taille de fragments suite à une restriction enzymatique puis une électrophorèse sur agarose. Les SNP ou substitutions de bases sont considérés comme une vraie forme de polymorphisme car ils constituent des marqueurs moléculaires beaucoup plus informatifs que les « indels » (Rafalski, 2002).

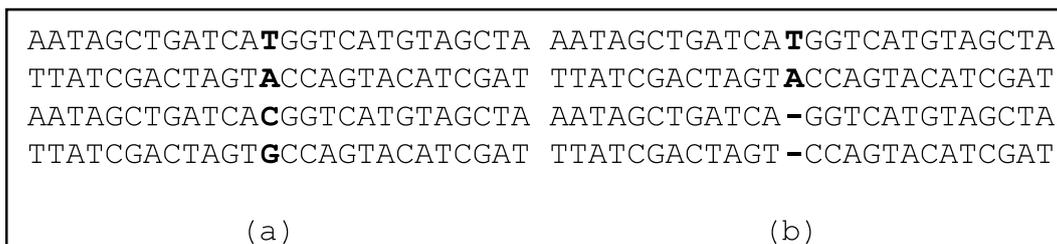


Figure 20 : Polymorphisme de séquence : (a) Polymorphisme de séquence d'un seul nucléotide (SNP) ; (b) Polymorphisme d'insertion ou de délétion (Indel).

3.3.2 La détection de SNP

Différentes approches ont été développées afin de révéler le polymorphisme SNP de l'ADN. En dépit de l'abondance des SNP au sein d'un génome, le développement et la caractérisation des SNP ne sont pas aisés.

3.3.2.1 La détection de SNP par approche « In silico »

L'une des méthodes principales de détection de SNP, est la méthode dite « *in silico* », rendu possible par l'augmentation constante du nombre de séquences disponibles dans les bases de données, notamment sur NCBI (www.ncbi.nlm.nih.gov/) et sur Sigenae (www.inra.fr/agenae/SIGENAE.html). Cette méthode consiste à rechercher des SNP dans les bases d'EST. En règle générale, les ESTs sont développés chez une espèce à partir de populations différentes, ce qui amène souvent à la redondance de certaines séquences et permet ainsi d'obtenir plusieurs séquences pour un même gène. Il est alors possible de détecter rapidement et facilement des SNP par analyse bioinformatique de telles séquences redondantes. La difficulté majeure de cette technique, réside dans le fait que les bases de données d'EST peuvent contenir des erreurs de séquençage qui doivent être distinguées du vrai polymorphisme. Actuellement, un grand nombre de logiciels sont capables de détecter de vrais SNP (revu par Marth et al., 1999). Dès lors que les SNP sont identifiés, l'étape d'amplification par PCR peut être effectuée en vue de caractériser le ou les SNP détectés. Il y a deux avantages principaux à cette méthode de détection des SNP. La recherche du polymorphisme est réalisée grâce à l'outil informatique à partir de données existantes, éliminant ainsi le coût du développement classique des SNP. Leur polymorphisme doit cependant être examiné pour le matériel biologique étudié (ex. les parents F0 pour leur cartographie).

3.3.2.2 Le séquençage direct

Cette seconde approche est plus directe. Elle consiste à détecter les SNP par le séquençage direct des produits de PCR issus d'un grand nombre d'individus à un même locus (Gaut et al., 1993). Des amorces sont dessinées à partir de séquences d'ADN connues extraites des bases de données. La recherche de SNP hors des régions codantes augmente considérablement la probabilité de détection de ces marqueurs moléculaires. Cependant, la recherche de SNP doit de faire préférentiellement au niveau de gènes présents en simple copie au sein du génome de l'espèce étudiée. En effet, si plusieurs copies du gène existent ou s'il existe une famille de gènes, plusieurs fragments d'ADN peuvent être amplifiés lors de la PCR. Ceci amène alors à d'importantes confusions dans la détection de SNP. Ensuite, si les individus génotypés sont homozygotes au locus considéré, la détection de polymorphisme est d'autant plus facile. Dans le cas d'individus hétérozygotes, les SNP peuvent être détectés mais les « indels » engendrent des difficultés. Les haplotypes ne peuvent pas être déterminés de façon exacte ou non ambiguë. Le clonage des produits de PCR peut lever cette incertitude en isolant chacun des haplotypes. L'autre solution consiste à croiser, par « backcross », l'haplotype incertain avec un individu dont l'haplotype est bien connu et de caractériser leur descendance.

La sélection d'amplicons dans des régions non codantes, tels que les introns, augmente la fréquence du polymorphisme observée d'un facteur trois environ (Zhu et collaborateurs, http://www.intl-pag.org/pag/9/abstracts/P3e_11.html). La PCR est alors basée sur un ensemble d'individus représentant la diversité de la population étudiée. Les produits de PCR sont alors séquencés dans les deux sens. Puis, la séquence résultante est alignée afin de distinguer les erreurs de séquençage du vrai polymorphisme.

3.3.2.3 Le polymorphisme de conformation de l'ADN

Ces techniques d'analyse électrophorétique des produits PCR sont particulièrement utiles pour mettre en évidence des mutations ponctuelles au niveau d'un gène. Des produits PCR (ou amplicons) particuliers sont formés si des mutations sont présentes sur l'ADN initial. Les produits PCR seront différents en fonction de la présence d'une séquence normale, d'une séquence mutée (homozygote, hétérozygote ou composite).

- Technique DGGE (pour « denaturing gel gradient electrophoresis »)

La température de fusion d'un produit PCR (ADN double brin), c'est-à-dire la température moyenne de séparation des deux brins est fonction de sa séquence. Une mutation ponctuelle qui change donc la séquence entraîne une modification de la température de fusion. Cette modification est mise en évidence par électrophorèse sur gel de polyacrylamide en présence d'un gradient d'agent dénaturant.

- Technique SSCP (pour « single strand chain polymorphism »)

La technique SSCP est basée sur l'analyse électrophorétique des produits PCR sous forme de fragments simple brin. On amplifie par PCR une région que l'on désire étudier et on compare la mobilité de l'ADN dénaturé portant une mutation par rapport à celle d'un fragment de référence comportant une séquence normale. Une mutation ponctuelle au sein d'une séquence modifie suffisamment la structure secondaire de l'ADN monobrin pour qu'il en résulte des changements de migration électrophorétique sur gel de polyacrylamide. Les techniques DGGE et SSCP sont concurrencées par des techniques de chromatographie liquide à haute performance avec des températures d'élution variables (DHPLC).

3.3.2.4 Le « Tilling »

Le Tilling, ou Targeting Induced Local Lesions in Genomes, est une technique de criblage des SNPs inconnus qui repose sur l'utilisation de l'endonucléase *Cell* isolée chez le céleri. Cette endonucléase clive les hétérohybrides d'ADN au niveau d'un mésappariement. Cette approche permet un criblage car la taille des fragments d'ADN utilisés est relativement longue (plusieurs kb). Après amplification par des amorces spécifiques du locus, des hétéroduplexes sauvages et variants sont générés par association, puis digérés par l'endonucléase *Cell*. Les produits de cette digestion sont alors visualisés sur gel dénaturant par électrophorèse (Till et al., 2004).

3.3.3 Le génotypage des SNP

Depuis les années 1990, un nombre croissant de techniques est désormais disponible pour génotyper des SNP en « haut débit » en plus de techniques plus traditionnelles. La plupart d'entre elles sont basées sur (1) le développement par amplification PCR puis (2) une réaction allèle – spécifique grâce à la connaissance exacte de la séquence d'ADN étudiée. Ces méthodes reposent principalement sur l'hybridation d'oligonucléotides, le clivage par des

nucléases, la ligation d'oligonucléotides, ou l'extension de primer (Gut, 2001). En voici une revue non exhaustive, présentant leurs principales caractéristiques (résumé en Figure 20). Chacune de ces techniques de détection de SNP présente des avantages et des inconvénients.

3.3.3.1 Restriction enzymatique

Si le SNP se situe sur un site de restriction d'une enzyme, il est alors possible de le détecter par le biais d'une PCR-RFLP. L'étude du profil des produits de PCR digérés sur gel d'agarose permet de révéler la nature du nucléotide au locus d'intérêt. Cette méthode ne nécessite aucun matériel spécifique qu'un laboratoire de biologie moléculaire ne posséderait. De plus, elle a l'avantage d'être peu coûteuse et relativement rapide mais le niveau d'information apporté reste faible. Bien que relativement fiable, cette technique n'est pas adaptée aux analyses à haut débit.

3.3.3.2 L'extension d'amorce

Cette technique nécessite l'utilisation d'une amorce spécifique de manière identique à la PCR. Il existe deux variantes de cette technique nécessitant un produit de PCR : dans le premier cas, deux oligonucléotides dont l'extrémité 3' de chacun d'eux est complémentaire aux deux allèles SNP sont utilisés. Ces deux oligonucléotides se différencient soit par leur taille soit par le fluorochrome qu'ils portent. Cette variante révèle alors chacun des deux allèles. L'autre variante consiste en une réaction de mini-séquençage aussi nommée Single Base Extension (**SBE**). Par le même principe, une réaction primer spécifique permet l'incorporation d'un des quatre dNTPs. Dans ce cas, ce sont les dNTPs qui seront différenciés par fluorescence, car chacun porte un fluorochrome différent. L'utilisation d'un lecteur de fluorescence, quel qu'il soit, permet de révéler l'allèle présent au locus étudié.

Plus récemment, la technique de **Matrice Assistée par Laser pour Désorber et Ioniser** l'échantillon, a été adaptée au génotypage de SNP. A l'origine, cette technique permettait d'identifier des protéines (Lester et al., 2002). Cette nouvelle variante est appelée **MALDI-TOF**. Tout d'abord, une réaction de PCR du type SBE (Single Base extension) est réalisée. Ensuite, l'échantillon est placé sur une plaque de métal pour être soumis à un rayonnement laser, qui propulse hors de la plaque le nucléotide incorporé au cours de la réaction de PCR. Celui-ci est alors accéléré à haut voltage pour voler ensuite dans un tunnel sans champ électrique, où est calculé son temps de passage (Time Of Flight). Ce temps de passage est converti en masse, en calculant le rapport de la masse (m) sur la charge du fragment (z) d'où (m/z) déterminant le temps de vol (ou **Time Of Flight**). Cette technique est, à l'heure actuelle,

considéré comme l'une des plus performantes en terme de répétabilité et de rapidité d'analyse (Gut, 2004). Elle permet également de déterminer la phase des allèles SNP au locus polymorphe c'est à dire de déterminer les bases qui se situent à ce locus chez un individu hétérozygote. La génération actuelle d'appareils de lecture MALDI-TOF est capable de lire un spectre toutes les secondes. En théorie, un débit maximum de 30.000 génotypages par jour peut être atteint, si les analyses sont multiplexées. Contrairement aux autres méthodes de génotypages SNP, la spectrométrie de masse repose sur l'analyse directe de la masse du produit et non sur une analyse indirecte où un réactif fluorescent ou radioactif est utilisé (Tost et al., 2002).

3.3.3.3 Le pyroséquençage

Le pyroséquençage est une technique récente de « re-séquençage » dont le principal avantage est sa rapidité autorisant ainsi l'analyse de grands nombres d'individus. Au cours de la réaction, les nucléotides ne sont pas ajoutés tous ensemble comme dans les réactions de séquence normales, mais l'un après l'autre. Si le nucléotide ajouté dans le milieu réactionnel correspond à celui attendu par la polymérase, il est incorporé dans le brin en cours de synthèse en libérant un PyroPhosphate. Grâce à une ATP sulfurylase, ce pyrophosphate est transformé en ATP, qui est utilisé par une luciférase pour émettre un signal lumineux. Ce signal est capté par le séquenceur qui le reproduit sous la forme d'un pic sur un pyrogramme. La hauteur de ce pic est fonction de l'intensité du signal lumineux, elle-même proportionnelle au nombre de nucléotides incorporés en même temps. La taille de la séquence est alors déduite des pics obtenus. Par ailleurs, en cas de mélange de nucléotides à une même position (individus hétérozygotes au locus), la taille des pics permet d'avoir une quantification de la proportion de brins porteurs de l'un ou l'autre des nucléotides (Ahmadian et al., 2000; Gruber et al., 2002).

3.3.3.4 La mesure du point de fusion par PCR en temps réel (rt-PCR)

Cette technique repose sur la détection et la quantification d'un marqueur fluorescent au cours de la réaction. Le signal fluorescent est directement proportionnel à la quantité de produits de PCR générés. En mesurant l'intensité de fluorescence émise à chaque cycle, il est possible de suivre la formation des produits pendant la phase exponentielle. Les produits d'amplification peuvent être détectés selon deux principes : (1) marquage non spécifique avec des agents se liant à l'ADN double brin (Sybr Green I[®]) ou (2) par marquage spécifique à l'aide de sondes fluorescentes ou sondes Taqman (AppliedBiosystems). La première technique a l'avantage d'être peu onéreuse. Pour distinguer les allèles, une courbe de fusion

est réalisée après l'étape de PCR. Les produits de PCR ont un T_m (ou température de fusion) différent selon l'allèle qu'ils portent, qui est clairement identifiable par la dérivée de la courbe précédemment établie (Schneeberger et al., 1995).

La seconde technique emploie deux approches au cours du marquage spécifique par sonde « Taqman » (figure 22).

- une sonde spécifique : la PCR est réalisée à l'aide de deux sondes encadrant le SNP et une sonde spécifique d'un allèle qui présente un double marquage situé en position 3' et 5', respectivement appelés *quencher* et *reporter*. Au cours du cycle, la fluorescence émise par le *reporter* est absorbée par le *quencher* situé à sa proximité. Pendant l'élongation du produit de PCR, la Taq polymérase va dégrader la sonde et le *quencher* va libérer sa luminosité. La fluorescence mesurée est proportionnelle à la quantité de molécules de produits de PCR contenant le SNP recherché (Livak et al., 1995).
- deux sondes spécifiques : dans ce cas, la PCR en temps réel nécessite l'utilisation de deux sondes Taqman spécifiques de l'allèle sauvage et du variant. Cependant, le principe reste le même. Il est alors possible de différencier un individu homozygote d'un individu hétérozygote au locus étudié, par le biais de la fluorescence émise.

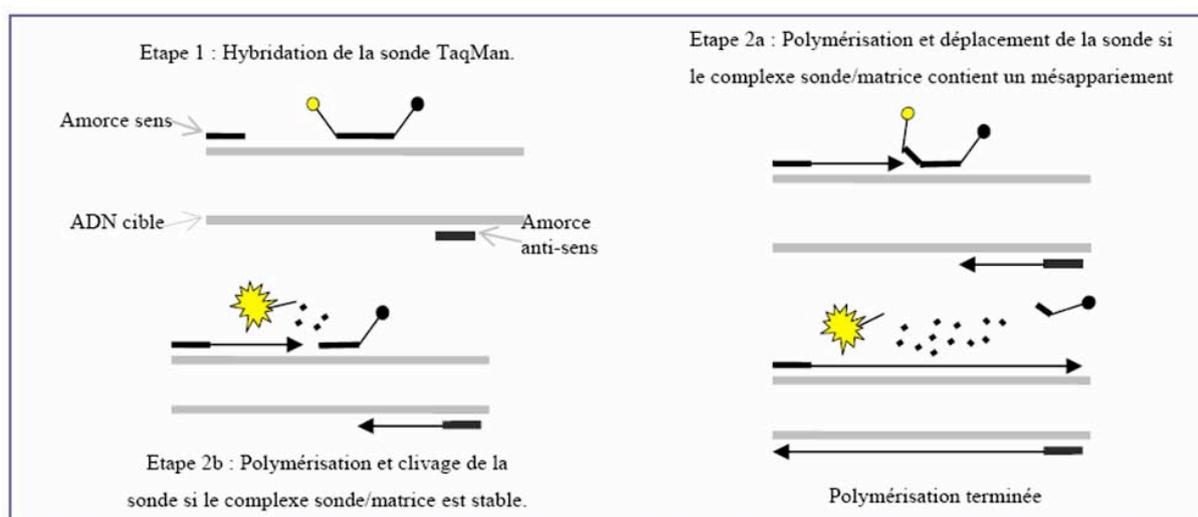


Figure 22 : Principe de la discrimination allélique par hybridation allèle spécifique de sondes TaqMan. Le fluorophore donneur est représenté en jaune et le fluorophore receveur en noir

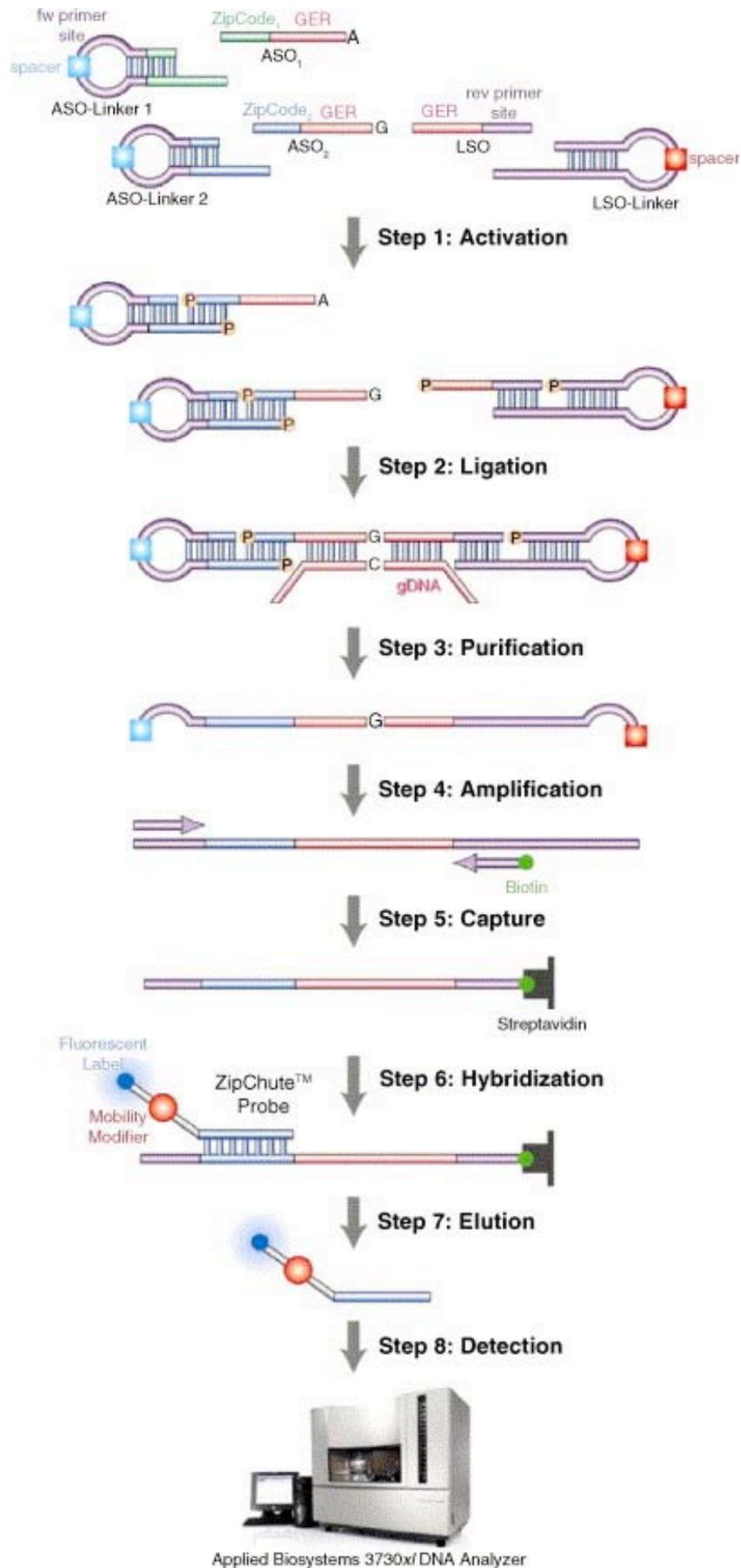


Figure 23 : Description des étapes du processus de génotypage par technologie SNPlex

3.3.3.5 Le SNPlex

Le système de génotypage de SNP, nommé SNPlex est une application de génotypage à haut débit qui utilise un ensemble de réactifs universels pré-optimisés. Ce système est relativement flexible car le génotypage des marqueurs se fait en multiplexe de 2 à 48 marqueurs par individu. Cela réduit ainsi le temps d'analyse et le coût du génotypage. Ce système de génotypage est basé sur la ligation d'oligonucléotides par PCR. Des sondes fluorescentes sont hybridées à des séquences complémentaires à la séquence spécifique de l'amplicon. Ces sondes sont détectées par électrophorèse sur séquenceur à capillaires. Cette approche est relativement intéressante par rapport aux technologies alternatives. Elle ne nécessite que trois amorces non marquées par marqueur SNP, consomme très peu d'ADN, et nécessite un séquenceur à capillaires qui est une machine que l'on retrouve désormais de plus en plus en laboratoire.

Le processus de génotypage se décompose en 8 étapes (figure 23)

1. Activation des sondes de ligation par phosphorylation
2. Ligation des 'linkers' sur les sondes complémentaires de la séquence d'ADN génomique
3. Elimination des oligonucléotides non ligués
4. Amplification simultanée des produits de ligation par PCR
5. Capture des amplicons
6. Hybridation de sondes universelles à l'amplicon capturé
7. Détachement des sondes universelles
8. Détection des sondes fluorescentes par électrophorèse

Le design expérimental du génotypage par SNPlex est généralement réalisé de façon automatique afin de déterminer la meilleure combinaison de sites polymorphes à génotyper et leur capacité à être génotyper en multiplex pour éviter toute interaction négative ou duplication. Le génotypage de SNP par la technologie SNPlex est une technique récente, relativement robuste, dont le coût est relativement faible puisqu'elle nécessite un équipement courant dans un grand nombre de laboratoires (De la Vega et al., 2005).

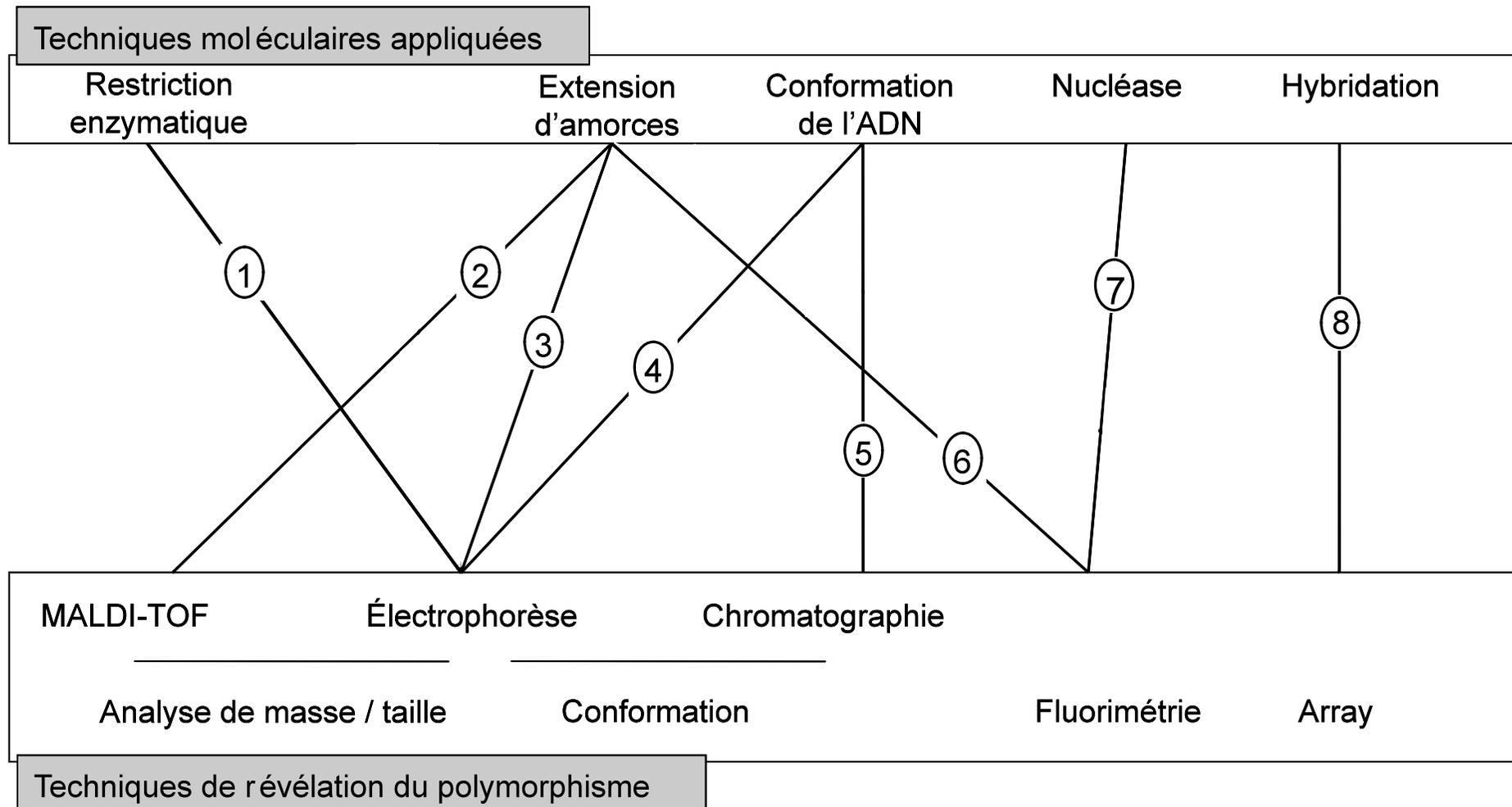


Figure 21 : Principales techniques de génotypage SNP (d'après Plomion (2005)

1 : PCR-RFLP ; 2 : Maldi-Tof ; 3 : Miniséquençage ; 4 : SSCP, DGGE ; 5 : DHPLC (Chromatographie) ; 6 : Pyroséquençage ; 7 : Taqman ; 8 : Microarray.

3.3.4 Applications des SNP

Au niveau mondial, le développement de marqueurs de type SNP date du milieu des années 1990. Depuis 2000, la base de données 'dbSNP : <http://www.ncbi.nlm.nih.gov/SNP>.' constituée par le « National Center for Biotechnology Information (NCBI) » réunit le plus grand catalogue de variations génétiques connues au sein de génomes divers (Sherry et al., 2001). A ce jour, cette base de données compte plus de 46 millions de sites polymorphes identifiés dont 16 millions ont été validés. La plupart ont été identifiés chez les principales espèces modèles telles que l'homme, la souris (*Mus musculus*) ou le riz (*Oriza sativa*). Le développement de marqueurs SNP est principalement lié au développement de séquences de type EST. En effet, cette ressource permet une sélection de sites polymorphes candidats pour leur intégration au sein de cartes génétiques (Picoult-Newberg et al., 1999 ; Kota et al., 2003). Chez les espèces d'intérêt aquacole, le développement de ce type de marqueur est initié. Par exemple, chez l'huître creuse américaine, *Crassostrea virginica*, des SNP ont été identifiés (Lee et al., 2006; Quilang et al., 2007) à partir de la production et de l'alignement de séquences d'ESTs. De même, chez trois espèces de saumon *Oncorhynchus tshawytscha*, *O. nerka* et *O. keta*, des SNP ont été développés avec succès à partir de séquences d'EST disponibles chez la truite arc en ciel *O. mykiss* et le saumon atlantique *Salmo solar* (Smith et al., 2005).

3.3.4.1 Cartographie génétique

Les marqueurs de type SNP sont utilisés dans de nombreuses applications telles que la construction de cartographie génétique de grande précision. L'une des approches proposées dans la cartographie génétique est de réaliser un « scan » complet du génome. Cela signifie qu'un nombre suffisant de marqueurs SNP sont développés et génotypés pour détecter des relations entre génotype et phénotype ou « QTLs » sur l'ensemble du génome. Cela a été réalisé chez le poulet notamment, où 2,8 millions de SNP ont été identifiés au cours d'un « genome Scan » (Ka-Shu Wong et al., 2004).

Sachidanandam et collaborateurs (2001) ont cartographié plus de 1,4 millions de ces marqueurs chez l'homme, avec une densité moyenne de un marqueur toutes les 1,9 kb. Cependant, chez les espèces d'intérêt aquacole, les seuls marqueurs SNP cartographiés à ce jour, ont été développés chez le Bar (Chistiakov et al., 2008).

3.3.4.2 Ancrage d'une cartographie physique

Il est possible de relier un groupe de liaison d'une cartographie génétique à une entité physique tel qu'un chromosome ou un gène. Cette carte physique est constituée de chromosomes BAC³ (Bacterial Artificial Chromosome) dont les séquences terminales (« Bac Ends ») sont relativement polymorphes. Ces BAC, assemblés en « Contigs », regroupent l'ensemble du génome de l'espèce considérée. Chez le maïs, approximativement 20% de ces « BAC ends » ont été séquencés et sont présents en une seule copie (Meyers et al., 2001). Ces séquences d'extrémités de BAC constituent de précieuses ressources pour la cartographie fine de région d'intérêt de génomes d'espèces modèles et le développement de nouveaux marqueurs SNP « *In silico* ».

3.4 SNP vs Microsatellites

Les marqueurs moléculaires microsatellites et SNP sont souvent qualifiés de marqueurs de « nouvelle génération ». Ce sont, en effet, les marqueurs moléculaires les plus utilisés dans divers domaines tels que la génétique des populations ou la cartographie génétique. Il est donc intéressant de dresser un comparatif entre ces marqueurs.

3.4.1 Traçabilité d'individus, d'un produit et utilisation médico-légale

Fries et collaborateurs (2001), ont proposé d'utiliser une signature génétique afin de suivre des animaux dans une chaîne de production alimentaire. Dans cette étude, le résultat principal est le suivant : en estimant le taux d'allèles nuls et d'erreurs techniques, un jeu de 8 marqueurs moléculaires microsatellites apparaît suffisant afin de permettre une identification exacte d'un bovin. Une comparaison a alors été menée afin de déterminer le nombre minimal de SNP nécessaire pour effectuer cette identification. A l'aide de marqueurs moléculaires bi-alléliques tels que les SNP distribués de manière aléatoire, le nombre minimum de SNP nécessaire a été estimé à 30. Ce résultat s'explique par la présence d'allèles rares au loci polymorphes (SanCristobal et al., 2000).

3.4.2 Assignation de parenté

En génétique, il est nécessaire de connaître le lien de parenté qui relie deux ou plusieurs individus (frère, sœur, demi-frère, parent, ...) entre eux. Le plan de croisement, le nombre d'individus impliqués, le taux d'erreur de génotypage, le nombre d'allèles au locus

³ Vecteur utilisé pour cloner des fragments d'ADN jusqu'à une taille de 300 kb.

sont les principaux facteurs qui influencent directement le pourcentage d'assignation de parenté (Letcher et al., 2001). Les travaux de simulations de Anderson (2006), tenant compte des erreurs de génotypages, montrent qu'un jeu de 60 à 100 marqueurs SNP lève toute ambiguïté sur une assignation de parenté quelle que soit l'espèce considérée. De plus, il a été estimé, avec des fréquences égales, qu'il faut en moyenne 2,23 fois plus de marqueurs bi-alléliques (SNP) que de marqueurs multi-alléliques (microsatellites) pour obtenir un même pourcentage d'assignation (Vignal et al., 2002). L'utilisation des microsatellites est donc plus facile dans ce cas, car l'effort de développement d'un nombre supérieur de SNP représente un travail très important.

3.4.3 Cartographie génétique et recherche fine de QTLs

L'utilisation de microsatellites pour la cartographie génétique est courante. Les premières cartes de liaison basées sur de tels marqueurs ont été établies chez nombreuses espèces d'intérêt aquacole comme le bar *Dicentrarchus labrax* (Chistiakov et al., 2005), la daurade *Sparus aurata* (Franch et al., 2006) ou encore l'ormeau *Haliotis rubra* (Baranski et al., 2006). L'inclusion de marqueurs moléculaires de type 1 (gènes) au sein de cartes génétiques n'est possible qu'au travers du développement de SNP. Cela a pour but de travailler en comparaison de cartographies chez des espèces proches ou de positionner des gènes d'intérêt. Quelques microsatellites ont été identifiés dans des zones très proches à des gènes (Gène A de l'amylase chez l'huître creuse ; Sellos et al., 2003) mais ce cas reste rare. Les SNP constituent donc des marqueurs de choix en cartographie génétique grâce à leur abondance et leur couverture quasi-complète des génomes (Sachidanandam et al., 2001).

Pour la recherche fine de QTLs, le polymorphisme responsable de la variance phénotypique se situe souvent au niveau de sites particuliers tels que les promoteurs (Vignal et al., 2002). Les marqueurs SNP révèlent alors leur intérêt en se situant au plus près des sites polymorphes qui conditionnent la réponse phénotypique. En effet, il a été démontré que l'utilisation de SNP pour la localisation de loci impliqués dans une maladie permet de réduire par 2 les intervalles de confiance sur la position de ces loci par rapport à l'utilisation d'autres types de marqueurs (Papachristou et al., 2006). De plus, les syndromes complexes sont souvent induits par de multiples loci dont la contribution de chacun dans la variance du phénotype est relativement faible (Papachristou et al., 2006). Cela souligne la nécessité de développer un grand nombre de marqueurs afin de localiser le plus finement possible les loci impliqués. Cependant, l'inconvénient des SNP provient de leur nature bi-allélique et leur faible taux d'hétérozygotie ($H < 0,5$). Cela confirme qu'ils sont moins informatifs que les

marqueurs microsatellites. L'hétérozygotie des marqueurs joue un rôle très important dans l'estimation des distances entre marqueurs. Les simulations menées par Leal et collaborateurs (2003) ont démontré l'intérêt d'utiliser des marqueurs microsatellites en association avec les SNP, car les cartes assemblées exclusivement d'un des deux types de marqueurs sont moins précises que celles assemblées à partir de l'association SNP-microsatellite.

3.5 Méthodes de développement et génotypage des marqueurs

3.5.1 Extraction de l'ADN

Au cours de ces travaux de thèse, l'ensemble des extractions d'ADN a été réalisé selon le protocole du Kit Qiagen (QIAamp DNA Mini Kit). La quantité d'ADN de chaque échantillon a été mesurée par spectrophotométrie (Eppendorf) puis diluer au 1/20^{ième} (100ng.μL⁻¹). En parallèle au dosage de l'ADN, la masse de tissu prélever a été mesurée au dixième de mg près afin de calculer le ratio masse de tissu/concentration en ADN. Des tests de validation par PCR ont été effectués afin de contrôler la qualité des extraits et standardiser les concentrations.

3.5.2 Développement de microsatellites *In silico*

En parallèle à l'ensemble de ces travaux, un effort de développement de nouveaux marqueurs microsatellites a été entrepris par une approche dite «*In Silico*» ou bioinformatique. Devant l'augmentation constante du nombre de séquences disponibles chez *Crassostrea gigas*, nous avons recherché des motifs d'ADN répétés (de 2 à 6 nucléotides) parmi les séquences d'EST obtenues au cours du Réseau d'Excellence Européen Marine Genomics Europe (<http://www.marine-genomics-europe.org/>; Tanguy et al., 2008). Les séquences dont le score était supérieur à 20 ont été retenues. Ce score est attribué par le moteur de recherche qui identifie les séquences répétées. Plus le score est important, plus la séquence répétée sera facilement utilisable (dessin d'amorces, taille de la séquence répétée, etc...). Dès lors, pour chaque marqueur, un couple d'amorces a été dessiné avec Primer 3 (Rozen et al., 2000) pour amplification sur une population naturelle composée de 48 individus. Pour plus de détails, l'article intitulé «*Identification and characterization of 18 novel polymorphic microsatellite markers derived from expressed sequence tags in the Pacific oyster Crassostrea gigas.*», présenté dans la section résultat, apporte un complément d'information concernant le développement de ces marqueurs.

3.5.3 Génotypage des marqueurs microsatellites

Plus de 300 marqueurs moléculaires microsatellites sont aujourd'hui disponibles dans la littérature et les bases de données (NCBI) pour *C. gigas*. En 2003, Li et collaborateurs (2003) ont développé et publié environ 80 marqueurs, qui, par la suite, ont été cartographiés par Hubert et Hedgecock (2004). Cette cartographie regroupant les marqueurs publiés par Huvet et al. (2000); Li et al. (2003) et Sekino et al. (2003), au sein des cartes génétiques relative à chaque sexe, est une base très utile nous permettant de sélectionner un ensemble de marqueurs selon leur emplacement. En effet, sur les cartographies génétiques développées par Hubert et collaborateurs, certains marqueurs sont relativement proches ($d < 10$ cM). En sélectionnant un ensemble de marqueurs dont la répartition est la plus large possible, nous avons cherché à optimiser l'établissement de la carte génétique. Cependant, une phase de travail préliminaire fut entreprise. Celle-ci avait pour but de vérifier les conditions de PCR nécessaires à la bonne amplification et d'identifier le niveau de polymorphisme de chacun des marqueurs publiés chez les 24 individus de la génération F0. Ainsi, les marqueurs très peu informatifs ont été écartés de la sélection.

3.5.4 Protocole expérimental

Les premières conditions de PCR testées correspondent aux conditions décrites dans la littérature. Les produits d'amplification ont ensuite été visualisés sur gel d'agarose à 1% ($40V.cm^{-1}$; 40 min). L'ADN utilisé est celui des 24 individus composant la génération F0 ($100ng.\mu L^{-1}$). Si un unique produit de PCR de taille attendue est observé, l'une des deux amorces marquées par un fluorochrome (Fam/Hex) adapté à un séquenceur à capillaires de type ABI 3100 a été commandée chez Eurogentec. Dans le cas contraire, si des bandes multiples apparaissaient, les conditions d'amplifications ont été modifiées par augmentation de la température d'accrochage des amorces (T_m), dans un premier temps. Si cela n'était pas satisfaisant, des gammes de concentration en amorces et en Chlorure de Magnésium ($MgCl_2$) ont été préparées, afin d'établir les conditions optimales d'amplification. Néanmoins, si aucun produit où un profil de type « multi bande » était obtenu, le marqueur moléculaire a été définitivement écarté de toutes les analyses suivantes.

Le mix réactionnel est commun à l'ensemble des marqueurs. Il se compose de 5 μL de Tampon, 5 μL de dNTP (2 mM), 1,5 μL de $MgCl_2$, 1 μL de chacune des deux amorces (12,5 μM), 1,5 μL d'ADN ($100ng.\mu L^{-1}$) et 0,2 unité de Taq polymérase (New England Biolabs). Le protocole d'amplification est le suivant : dénaturation initiale de 5 min à 96°C, puis 30 cycles

de 1 min à 96°C, 30 sec au T_m optimal, puis une élongation de 1 min à 72°C et enfin, une phase d'élongation finale de 30 min à 72°C.

La taille des allèles microsatellites est lue après migration sur un séquenceur à capillaires (ABI3130, Applied Biosystems). Dans un souci d'économie et de rapidité, les produits de PCR d'un même individu ont été groupés par deux selon le fluorochrome utilisé au cours de la PCR. Cela permet de faire migrer au cours d'un même run d'analyse les produits de deux marqueurs. Les allèles de ces derniers sont différenciés par la couleur du fluorochrome porté par les amorces de PCR (lecture deux à deux, fluorochrome Fam couplé avec un fluorochrome Hex). Cette stratégie a été appliquée à l'ensemble des marqueurs microsatellites analysés. Un volume de 1µL de chacun des deux produits de PCR est ajouté à un mix composé de 0,25µL de marqueur de taille (ROX GS500, Applied biosystems) et de 7,75µl de Hi-Di formamide (Applied biosystems). L'ensemble est dénaturé pendant 5 minutes à 96°C puis rapidement placé sur glace (4°C) pour éviter la renaturation des deux brins d'ADN. Ensuite, les échantillons sont placés dans le séquenceur. Les paramètres d'analyse de la machine sont conformes aux recommandations du constructeur (temps d'injection de 100 ms, voltage de 20 kV, temps de migration de 2500 sec).

La lecture des profils de migration (ou électrophérogramme) est un travail long et fastidieux qui nécessite de vérifier notamment le bon positionnement du marqueur de taille sur lequel repose l'estimation des tailles des allèles. Un examen visuel de ces profils permet de vérifier que les pics, retenus par le logiciel, correspondent bien à un allèle microsatellite et non à un produit non spécifique se situant à proximité. Le logiciel utilisé est bande® v4 (Applied Biosystems) dont les paramètres d'analyse par défaut ont été respectés.

3.5.5 La catégorisation des groupes d'allèles ou « Binning »

Afin de caractériser et de corriger les groupes d'allèles, c'est-à-dire de ne retenir que les allèles parentaux qui apparaissent dans le génotype des descendants, l'étape dite de «Binning» est indispensable. Le logiciel (Genemapper®) permet par l'élaboration de «panels» (définition d'une fenêtre d'apparition des valeurs alléliques de descendants pour chaque locus en fonction des allèles parentaux), de déterminer et de valider les allèles réellement transmis tout en s'affranchissant des pics artefactuels et de la variabilité entre migration. Le binning est une étape importante car la qualité des données en dépend entièrement. Cette étape est partiellement automatisée par l'utilisation d'un logiciel dédié mais nécessite une vérification manuelle et visuelle des allèles amplifiés, la présence de pics parasites entraîne parfois des

confusions dans la lecture des génotypes. Pour éviter ce type d'erreur, l'ADN des grands-parents (F0) et des parents (F1) a été systématiquement placé sur la microplaque de PCR. Pour un même « run » d'analyse, il est possible de comparer directement les allèles parentaux et les allèles des descendants et d'écarter les pics parasites.

3.5.6 Développement de marqueur SNP

Au début de l'année 2008, la base de données Genbank contient plus de 22000 séquences EST principalement issues de travaux d'expression de gènes chez *C. gigas*. Le nombre d'EST disponibles dans le domaine public va atteindre plus de 50000 séquences grâce au dépôt des EST séquences par le Génoscope et le projet Marine Genomics Europe. Cependant, fin 2004, les séquences EST disponibles dans le domaine public, se limitaient à 4428 EST déposés sur la GenBank et 1142 autres séquences du même type déposées dans la Gigasbase (Gueguen et al., 2003; <http://www.ifremer.fr/GigasBase/>). La recherche de polymorphisme nucléotidique a été orientée vers des EST de fonctions connues. Contrairement à des EST anonymes, cela présente l'avantage de cartographier des gènes de fonctions connues. Dans l'éventualité où ces marqueurs co-localisent avec des régions QTL, il est alors intéressant de mettre en relation la fonction du gène et sa potentielle implication dans la variance du caractère considéré. Par exemple, la recherche de polymorphisme a été effectuée, en partie, sur des EST hébergés sur la Gigasbase, dont l'ensemble des séquences correspond à des gènes impliqués dans des fonctions immunitaires.

Pour chacun des EST sélectionnés, un premier couple d'amorces a été dessiné par l'utilisation de Primer 3 (Rozen et al., 2000) pour obtenir des amplicons de taille comprise entre 350 et 550 paires de bases. Puis, 3 conditions de PCR différentes ont été testées pour déterminer les conditions optimales d'amplification. Ces conditions diffèrent par leur température d'annealing. Pour la première condition de PCR, dite en 'touchdown', la température d'annealing couvre une gamme allant de 55 à 45°C, la seconde couvre la gamme de 65 à 55°C et la troisième couvre la gamme de 74 à 65°C. Cette stratégie d'amplification par PCR en 'touchdown' augmente la probabilité d'amplification des fragments. En effet, en appliquant une température d'annealing qui décroît de 1°C tous les 2 cycles, cela augmente la spécificité de la réaction. Lorsque la température d'annealing atteint la température maximale d'annealing du couple d'amorces, les premières copies d'ADN produites sont très spécifiques. Dès lors, le différentiel entre le nombre de copies de l'amplicon et le nombre de copies d'ADN « parasite » ou non attendu penche largement en faveur de l'amplicon. Généralement, le signal obtenu par visualisation du produit de PCR sur gel d'agarose est plus intense lorsque

celui-ci a été amplifié par 'touchdown' PCR. Cette stratégie permet également l'amplification de différents produits dont la température d'annealing des amorces est relativement proche ($\pm 4^{\circ}\text{C}$) au cours d'une même réaction de PCR. Ainsi, les couples d'amorces dont les températures d'annealing étaient identiques ou proches ont été groupés par nombre de 4 (4×24 individus $F_0 = 96$, c'est à dire le nombre exact de puits disponibles sur une plaque de PCR). Le protocole appliqué est le suivant :

Un mélange ou « mix » commun a été composé pour l'ensemble de la phase de recherche de polymorphisme. Il se composait de : 5 μL de dNTP (2 mM - Eurogentec), 5 μL de Buffer (1X), 1 μL de MgCl_2 (2 mM - Sigma), 1 μL de chacune des deux amorces (10 μM - Eurogentec), 0,3 unité de Taq polymérase (New England Biolabs), 2 μL d'ADN ($100\text{ng} \cdot \mu\text{L}^{-1}$) q.s.p 48 μL d'eau de biologie moléculaire (Sigma). Les conditions de PCR sont également communes : tout d'abord une phase de dénaturation de 5 min à 96°C est appliquée, puis pour chaque palier de 1°C au cours de la descente de la température d'annealing, 2 cycles composés d'une dénaturation de 30s à 96°C , d'une phase d'annealing de 30s au T_m et suivi par une phase d'élongation de 1 min à 72°C . Enfin, lorsque la température d'annealing basse est atteinte, 25 cycles d'amplification sont appliqués identiques au précédent et à cette même température basse (45/55 ou 65°C).

L'étape suivante consiste à purifier une fraction des produits de PCR afin d'éliminer tout élément qui interfère au cours de la réaction de séquençage. Il s'agit principalement d'éliminer les dNTP et les amorces non incorporées au cours de la réaction de PCR. En effet, au cours de la réaction de séquence, ceux-ci vont jouer le rôle de brins matrices et permettre la fixation de fluorochromes contenus dans le Mix réactionnel. Ces produits vont alors se mêler à la séquence cible en cours de formation et sa lecture sur séquenceur sera problématique voir impossible. Sur les électrophérogrammes, chaque nucléotide de la séquence cible présente deux pics différents perturbant l'assignation des bases par le logiciel d'analyse de séquences (Seqscape v2.1, Applied Biosystems). De telles séquences ne sont pas exploitables. Pour se faire, l'enzyme Exo-Sap IT (Amersham Biosciences), isolée chez la crevette, est utilisée. Elle a la propriété de dégrader les petits fragments d'ADN et d'être inactivée par une exposition à une température de 80°C pendant plusieurs minutes. Un volume de 2 μL de cette enzyme permet la purification d'un volume de 5 μL de produits de PCR lorsque qu'elle est placée en mélange et à 37°C pendant 30 min. Ensuite, elle est inactivée afin de ne pas dégrader les produits de PCR par une exposition à 80°C pendant au moins 15 min. Cette enzyme n'est pas la seule méthode de purification de produits de PCR. Il est également possible de réaliser cette étape par précipitation à l'acétate d'ammonium ou par centrifugation sur micro-colonne

(Microcon - Eppendorf). Cependant, l'utilisation de cette enzyme est la solution la plus rapide (45 min suffisent pour purifier 96 échantillons) et son coût est relativement faible (0,5€/échantillon).

L'étape suivante consiste à réaliser la réaction de séquence sur les produits de PCR purifiés. Le Mix réactionnel (Big Dye Terminator® v3, Applied Biosystems) a été optimisé dans le but de réduire le volume de réactifs nécessaires au cours de la réaction par rapport au protocole recommandé par Applied Biosystems. Ainsi, le volume réactionnel a été réduit de 25µL (recommandé) à 10µL. Ce dernier se compose de 2 µL de Tampon 5X, de 0,5µL de Big Dye terminator® v3 (mix de dNTP portant un fluorochrome spécifique à chacun des 4 nucléotides de l'ADN), de 1 µL de l'une des deux amorces ayant servi à l'étape précédente de PCR (5 µM), d'un microlitre de produits de PCR purifié (100 à 300 ng.µL⁻¹) et enfin d'eau de biologie moléculaire q.s.p 10 µL. Le cycle d'amplification est toujours le même quel que soit le produit à séquencer. Il se compose d'une dénaturation initiale de 5 min à 96°C suivie par 25 cycles de 10 s de dénaturation à 96°C, 5s d'annealing à 55°C et d'une élongation de 4 min à 60°C.

Ensuite, la dernière étape consiste à éliminer de nouveau les produits non incorporés au cours de la réaction de séquence (amorces et dNTPs marqués). De même, il existe plusieurs méthodes de purification dont l'efficacité et le coût est relativement variable. La précipitation par éthanol est l'une des solutions les plus efficaces et surtout la moins chère. Le protocole appliqué est le suivant :

- (1) Ajouter aux 10 µL de produits de réaction de séquence, 26 µL d'eau ultra pure (Sigma) et 64 µL d'éthanol (>95%) fraîchement préparé.
- (2) Centrifuger à 3000g pendant une durée de 30 min
- (3) Vider les tubes puis laver avec 100 µL d'éthanol à 70%
- (4) Centrifuger à 3000g pendant une durée de 15 min
- (5) Vider les tubes puis sécher au SpeedVac pendant 15 min
- (6) Resuspendre les culots secs avec 10 µL de formamide Hi-Di (Applied Biosystems) pendant 30 min à l'abri de la lumière

Enfin, les échantillons sont dénaturés une dernière fois (96°C, 5 min) puis placés sur glace. Les propriétés dénaturantes du formamide ainsi que le passage rapide des échantillons

de 96°C à 4°C (flash-cooling), évite la ré-association des deux brins complémentaires de l'ADN car les séquences doivent être lues sous forme mono-brin. Ces séquences sont placées dans un séquenceur à capillaires (ABI 3130, Applied Biosystems). Les paramètres d'analyse de la machine sont conformes aux recommandations du constructeur (temps d'injection de 100 ms, voltage de 20 kV, temps de migration de 2500s).

Les profils de migration bruts (ou électrophérogrammes) sont analysés grâce au logiciel Seqscape v2.1 (Applied Biosystems). Ce logiciel permet à la fois de révéler la séquence lue sous forme d'un enchaînement de nucléotides à partir des profils bruts (raw data) mais aussi d'aligner les séquences les unes par rapport aux autres. Dès lors, le travail de nettoyage et de correction manuelle des erreurs de lecture des séquences commence. Cette étape a pour but de corriger (si nécessaire) l'alignement des séquences formé par le logiciel et de lever les incertitudes de l'assignation des bases. Lorsque deux pics se superposent sur la même position, par défaut, le logiciel considère le nucléotide comme un site indéterminé (noté N). Par examen de la qualité de la séquence (QV : quality value) et de la hauteur relative de chacun des pics qui se superposent, l'incertitude est levée. Le site polymorphe est soit confirmé, soit modifié. Ce travail est relativement long mais il est nécessaire et très important. La présence de faux sites polymorphes tels que les singletons est à éviter. Tous les singletons identifiés ont été ignorés si chacun de ces sites polymorphes n'était pas présent chez au moins deux individus frère-sœur ou sœur-sœur.

Chaque alignement, constitué des 24 séquences d'un même gène chez les individus de la F0, est transformé au format Fasta (extension .fasta). Ce format, largement répandu, permet l'exploitation des données par d'autres logiciels plus spécialisés. Pour chacun des alignements, le cadre de lecture (ou ORF - Open Reading Frame) a été déterminé afin de caractériser les sites polymorphes par l'utilisation des logiciels Bioedit (Hall, 1997) et DnaSP (Rozas et al., 1995). Pour chaque site polymorphe, les informations suivantes ont été déduites :

- Position en paires de bases ; marqueur situé dans un intron, un exon ou un UTR
- Nature : Transition (Ts), Transversion (Tv), synonyme (pas de changement d'acide aminé), non-synonyme, silencieux et non sens (introduit un codon stop)
- Diversité nucléotidique de la séquence (II)

3.5.7 Analyse du biais d'usage du code génétique chez *C. gigas*

Afin d'étudier le biais d'usage du code génétique chez *C. gigas*, le tableau des codons optimaux a été déduit à partir de l'analyse d'un jeu de séquences EST. La redondance de chaque EST a été utilisée comme une évaluation approximative de l'expression des gènes (Duret et Mouchiroud, 1999). Le jeu de données est composé de 8800 séquences EST issues de bande (<http://www.ncbi.nlm.nih.gov/>) et des données obtenues au cours du réseau d'excellence Marine Genomics Europe. Les cadres de lecture (ORFs – Open reading frames) ont été identifiés et les séquences d'une longueur inférieure à 100 codons (300 bp) ont été retirées du jeu de données pour éviter d'utiliser de mauvais cadres de lecture et avoir assez de codons pour calculer la fréquence des codons synonymes. En effet, plus un cadre de lecture est long, plus la probabilité qu'il soit correct est élevée. Il est cependant possible que certaines annotations d'EST soient fausses à la suite de substitutions dans la séquence nucléotidique. Nous avons supposé que les problèmes de qualité des séquences n'introduisaient pas de biais d'analyse.

Ensuite, à l'aide de l'algorithme développé par Bazin et collaborateurs (2005), nous avons regroupé toutes les séquences appartenant à un même gène pour construire la base de données. Deux séquences ont été regroupées si elles partagent un fragment de plus de 300 pb d'une similitude supérieure à 95%. Seule le plus long cadre de lecture de chaque regroupement de séquences a été utilisée pour calculer le tableau d'usage des codons. La fréquence des codons synonymes par acide aminé a été calculée avec le programme CODONW (Peden, 1999). Un synonyme codon a été qualifié d'optimal lorsque sa fréquence augmente de façon significative avec la redondance des EST (Duret et Mouchiroud, 1999). Cette corrélation a été testée par le test non-paramétrique de la corrélation de Spearman (JMP v5.0© ; SAS Institute Inc.). Dès que les codons optimaux ont été identifiés, nous avons utilisé la fréquence des codons optimaux (Fop; Ikemura, 1985) comme la mesure du biais d'usage du code génétique chez *C. gigas*.

3.6 Critères de choix des loci polymorphes et leur génotypage chez les individus F2

Parmi l'ensemble des loci polymorphes identifiés dans une même séquence d'EST, un seul site polymorphe a été choisi pour être génotypé par SNPlex. Plusieurs critères ont été établis afin de retenir le site le plus approprié. Tout d'abord, le site polymorphe doit être relativement isolé d'un autre site polymorphe pour autoriser le dessin d'une amorce spécifique. Cet intervalle a été fixé à 20 paires de bases en amont et en aval du SNP et correspond aux exigences de la technologie du SNPlex. Le second critère est celui du niveau d'information apporté par le marqueur. Cette information est cruciale car l'analyse en ségrégation en dépend totalement. Nous avons donc préféré choisir des sites pour lesquels l'un des deux grand-parents de chaque couple de F0 est homozygote pour le marqueur et l'autre grand-parent hétérozygote à ce même locus (Figure 24).

Grand-Parent 1	Grand-Parent 2
-TCAC C GGT-	-TCA T GGT-
-TCAC C GGT-	-TCA T GGT-

Figure 24 : Cas idéal de marqueur SNP pleinement informatif (pas d'ambiguïté sur l'origine parentale de l'allèle C).

Cette stratégie a été appliquée à l'ensemble des séquences EST pour identifier le site polymorphe le plus intéressant pour le plus de familles possibles. Le travail conséquent de génotypage des marqueurs de type SNP chez les 900 individus de la F2, a été réalisé sur la plate forme du Max Planck Institute (MPI) à Berlin (Allemagne). Ce laboratoire a été désigné comme plateforme partenaire du réseau d'excellence « Marine Genomics Europe ». L'ADN des individus de la F2 ($n=900$), de la F1 ($n=10$) et de la F0 ($n=20$) a été envoyé sous format de plaques PCR de type 96 puits à une concentration de $10\text{ng}\cdot\mu\text{L}^{-1}$ ($V=50\mu\text{L}$). Pour plus de sécurité dans le choix des sites polymorphes à génotyper, nous avons choisi d'ajouter l'ADN des individus de la F0 et de la F1. Cela nous a permis de vérifier une seconde fois le génotype des individus aux loci retenus. Les résultats du génotypage ont été codés en données binaires (1/2 : individu codé 11 ou 22 = individu homozygote ; individu codé 21 ou 12 = individu hétérozygote).

3.7 Résultats

3.7.1 Microsatellites

Un total de 114 marqueurs, choisi pour leur couverture la plus large et régulière sur la carte de liaison et publiés par Hubert et al. (2004) ou publiés par Yamtich et al. (2005) et Yu et Li (2007) ont été testés sur un échantillon d'individus d'une population naturelle. Après migration des produits de PCR sur gel d'agarose (1,5%) puis sur séquenceur à capillaires, les marqueurs montrant une seule bande suffisamment intense sur gel d'agarose ont été retenus. 65 marqueurs ont été écartés pour les raisons suivantes : aucune condition de PCR testées n'a permis d'amplification pour 16 marqueurs ; 30 marqueurs montraient systématiquement plus de deux allèles par individu ; les 19 marqueurs étaient monomorphes. Un total de quarante neuf marqueurs microsatellites (soit 43% - Tableau 11) ont été retenus et génotypés chez l'ensemble des individus étudiés (soit 20 F0, 10 F1 et 900 F2). Le nombre moyen d'allèles par marqueur est 7,3 avec au maximum 14 (*UcdCg* 126,156, 165 et 196) et au minimum 2 (*Cgri* 3, 4, 45, et *UcdCg* 145, 152, 176 et 193).

Les données de génotypes après « binning », corrections et de vérifications manuelles des allèles ont été déposées sur la base de données Respecies abritée par le Roslin Institute (<http://www.respecies.org/>). Les données ont été codées selon la taille de chacun des allèles pour chaque marqueur.

Tableau 11 : Caractéristiques des marqueurs microsatellites génotypés

Locus	Séquence Répétée	Séquence des Amorces (5' - 3')	Ta (°C)	Na	Gamme de taille (pb)	Numéro d'accèsion GenBank	Publications associées
Crgi3	(TC)n	TAGGATGAGGCTGGCACCTTGGAGCCTGCCTTGCCTTTGAGGAATA		2	166-168	AB091548	(Sekino et al., 2003)
Crgi4	(TTC)n(CTTTT)n	CCAAAACACGATAAGATACACTTTCGATCAGTCCCTCACATCTTTCCTC		2	243-245	AB091549	(Sekino et al., 2003)
Crgi26	(TC)n	ACAACCCCAATAACATTAGGAAGATATTGTAACACTGGATGAATCTGGG		9	177-209	AB091570	(Sekino et al., 2003)
Crgi45	(AT)n(ATTT)n	GAGTCACCATGAAGAGTATCTGAAATGATTACATAACTCTGACCCAAT		2	147-159	AB091588	(Sekino et al., 2003)
UcdCg109	(CAT)n	GCTATGGTTGTCATCCTCGAATGCCTTTATCGGTTTTGCTT		10	164-225	AF468525	(Li et al., 2003)
UcdCg117	(TC)n	CCAAGCTTGCCTCACTCAAAGAGTGTCTGGTGTGCCAAAT		8	288-332	AF468528	(Li et al., 2003)
UcdCg119	(TC)n	AGGATGCCAATCGATTTTATTTACCATGCCGTCTTAGTGGAC		5	208-222	AF468529	(Li et al., 2003)
UcdCg120	(CA)n (GA)n	GGGTGAGATTTAGGGGGGAGACTCCATCAAACCTGCCAAAC		4	149-155	AF468530	(Li et al., 2003)
UcdCg126	(TCTA)n	TGGATTTGATCACCCCTTACACCTGGATTCTGTGCGAGATT		14	116-207	AF468532	(Li et al., 2003)
UcdCg129	(GA)n	CGAATTTTTTCGGACATCGTTGTGGTATGCCTGCATCATGT		13	222-255	AF468534	(Li et al., 2003)
UcdCg133	(CT)n	GTTGGCGCGAAATATAGGAAATGAACTTGTGATGTGCAGGA		5	278-312	AF468537	(Li et al., 2003)
UcdCg134	(CT)n (CGCT)n	TTTGTATGTCCGTCATCGTCAACCTGCAAAATGGTGAAAAT		5	204-244	AF468538	(Li et al., 2003)
UcdCg141	(TG)n	CTCAACGACTTTTTGCCTGAGTGTCTTCTATCCCGCAACG		11	227-262	AF468545	(Li et al., 2003)
UcdCg145	(GA)n	CAGGTGTATGCGACGTGACTCAGGCTTTTAAGCGCATTTT		2	255-258	AF468547	(Li et al., 2003)
UcdCg149	(GA)n (GACA)n	TGATTAACGTGGGTGATTGAGTTTCTGACTGTCCGTCTGTGA		7	213-245	AF468551	(Li et al., 2003)
UcdCg150	(CT)n	CCTTTCATGTAGGATGACTCTTTTTTGAGGGGAGCATTGATTGAT		8	250-291	AF468552	(Li et al., 2003)

Locus	Séquence Répétée	Séquence des Amorces (5' - 3')	Ta (°C)	Na	Gamme de taille (pb)	Numéro d'accèsion GenBank	Publications associées
UcdCg151	(GT)n	AGGTAATCCGCAAACCCAGTG GCATTGCGTCAGGATTAGGT		8	264-308	AF468553	(Li et al., 2003)
UcdCg152	(CAT)n	TGGTTTTGGAGCTTGGCTTA TCAAGCAAAGAAAGTCACCTCA		2	256-259	AF468554	(Li et al., 2003)
UcdCg153	(GAT)n	GCAGCAGCTTCAGAGTGAAA CAAAATCTGGTGGACCTTCG		11	257-289	AF468555	(Li et al., 2003)
UcdCg156	(GATA)n (GA-TG)n (TA)n	AGCAGACCTTGGCAAATACG CCGTCATCAGGTCCTGTTTT		14	116-208	AF468557	(Li et al., 2003)
UcdCg157	(GA)n (TAGA)n	GGGGGATGTCCGAGAAGTAT AACAGAGAAAGGTGGATTTTAGGA		9	229-313	AF468558	(Li et al., 2003)
UcdCg160	(GA)n (GACA)n	GGAGCCATTAACAACACCACA TCTCTCCCTTCCCCCTCTTA		11	223-278	AF468560	(Li et al., 2003)
UcdCg162	(TTCA)n (AT-CT)n (GTCT)n	CCAAATCACCGTTTTAGTTTGT AGCGACACAGAGACCACCTT		6	199-326	AF468562	(Li et al., 2003)
UcdCg165	(ATCT)n (CA)n	TTTTTACCAGCACTCGCTGT TCCGAATTTACAAGTGTGTGT		14	171-239	AF468565	(Li et al., 2003)
UcdCg166	(TC)n	CATCGGAATAAATCGGGTAA TTCCTTTGTGCTGTCTTACAGG		6	204-237	AF468566	(Li et al., 2003)
UcdCg170	(GA)n (GT)n	TGGTGGTCAGTGAATGTGAGA CGGACAGTAGCCTTTTAACACA		8	110-134	AF468568	(Li et al., 2003)
UcdCg171	(CAT)n	CCACTCATAAGGGAAAATGAAA TCGTGACCTTAAAACCTCGT		8	245-262	AF468569	(Li et al., 2003)
UcdCg172	(GAT)n	CCACCGTTAAACGTAGCATTG TTGTGTCCCTTTTCCGCTC		5	242-250	AF468570	(Li et al., 2003)
UcdCg173	(CT)n (CA)n	AAAATGGGAATTCAGTGTGCA CGGCACCGTTTTGTTATCT		9	217-250	AF468571	(Li et al., 2003)
UcdCg174	(CAT)n	CTTCCTGCTGCAGAACCTGT AATGACGGGATGATGATGATG		5	178-244	AF468572	(Li et al., 2003)
UcdCg176	(GAT)n	TTCCGATGATGATAGCGATG GGCTCGTGTTCCAATATGGT		2	259-262	AF468574	(Li et al., 2003)
UcdCg180	(GAT)n (GTT)n	TCACACGCAGCGAATTTTAA AATAACCACGCCGACAGC		6	281-306	AF468578	(Li et al., 2003)

Locus	Séquence Répétée	Séquence des Amorces (5' - 3')	Ta (°C)	Na	Gamme de taille (pb)	Numéro d'accèsion GenBank	Publications associées
UcdCg181	(GT)n (GA)n	CACCCCAAAGGACCACATAC TGTCAGCATGGGTAAGTCCA		10	214-263	AF468579	(Li et al., 2003)
UcdCg186	(GA)n	GCCGCCGATTCTCTTAGATT GGGCTAGCTAGTCATCACCCCTA		6	243-284	AF468584	(Li et al., 2003)
UcdCg189	(GA)n	AATGCAGGGGTCACCTGAGAG GGGATTAAGTTTGAACATGTGG		8	159-262	AF468587	(Li et al., 2003)
UcdCg191	(GA)n (CAGA)n	GCTTCCATGACACAACACTACTGG ATGATGCATTGCCAAGTTGA		11	236-258	AF468589	(Li et al., 2003)
UcdCg193	(GT)n (TA)n	CTAGAAACGCTTCGGTTCGAT TATCTTTCGCAAATCGCAGT		2	252-254	AF468591	(Li et al., 2003)
UcdCg194	(GAT)n (GAG)n	CCCAGTGAAAACCTGGAGACA TTTCGAATCGGGAAAATACG		9	244-290	AF468592	(Li et al., 2003)
UcdCg195	(CAT)n	CCAACAACAGGGCACCTACT GGTCCAGTTGGCATTCTCTA		6	260-282	AF468593	(Li et al., 2003)
UcdCg196	(GAT)n (GAC)n	CCTTTCATTTGGAGTTACATTG ATCTTGCCATTTGCTTTTGG		14	245-318	AF468594	(Li et al., 2003)
UcdCg197	(CTT)n (TG-A)n (GA)n	AGCAGACCCACTGGAGGTAA GTCGCTTCACCCAGGAAAT		10	238-364	AF468595	(Li et al., 2003)
UcdCg198	(CAT)n	GAAAGACACGACCGGAGAGA CTGATGATGTCCCACACCTG		8	220-250	AF468596	(Li et al., 2003)
UcdCg199	(CAT)n	GGGAAGAGTTGAATTCTGCAA AAACCGAGGCTCAGGAAAAT		3	265-273	AF468597	(Li et al., 2003)
UcdCg200	(GAT)n	AAAGTTGCTTTGCTGTCGTC CGCTAACGTGCTTCATTCAA		11	232-265	AF468598	(Li et al., 2003)
UcdCg203	(GATA)n	AAGCTTACCGGACTGGGTTA GGGTTCCGCCATTTACTTTA		5	129-282	AF468601	(Li et al., 2003)
UcdCg210	(GAT)n	TTCACAATGAAGATGACAGTGC CCTCCTCTGCCTCCATATCA		4	318-325	AY999708	(Yamtich et al., 2005)
CGE001	(TG)n	TCGCTATACTGAGATAGACG GACTGTACTTCGCTGTTC		2	195-197	AJ565507	(Yu et al., 2007)
CGE006	(TA)n	TAACCTCTTCAAGGATTT ATGTCATTATCGCTCAA		2	313-324	BQ426914	(Yu et al., 2007)
CGE009	(AG)n	TTCGTTGAAGGTGACAAGTG GCATTTTGGGATGAACAGA		8	107-129	CX068958	(Yu et al., 2007)

3.7.2 Développement de marqueurs microsatellites par approche « *In silico* »

L'ensemble des résultats du développement de marqueurs microsatellites par approche *in silico* sont détaillés au sein d'un article présenté dans les pages suivantes (Sauvage et al., 2008). Cet article a été accepté pour publication dans le journal *Molecular Ecology* Ressources. L'encadré 1 résume les résultats obtenus pour le développement de ces marqueurs SNP.

Encadré 1: Synthèse des résultats du développement des marqueurs Microsatellites *in silico*

Les travaux de recherche de microsatellite ont permis d'obtenir les résultats suivants:

- 22 séquences répétées ont été identifiées par « screening » des séquences EST produites dans le cadre du Réseau d'excellence « Marine Genomics Europe » (Tanguy, 2008).
- 18 des 22 couples d'amorces ont donné un unique amplicon et ont été amplifiés chez 48 individus issus d'une population naturelle (Bassin de Marennes-Oléron, France)
- Le nombre moyen d'allèles est de 29 avec un minimum de 12 et un maximum de 48 allèles
- Le niveau d'hétérozygotie attendu se situe entre 0,77 et 0,97
- 3 marqueurs apparaissent liés entre eux (Sili3, Sili6 et Sili7; $P > 0.05$)
- 3 marqueurs dévient de l'équilibre d'Hardy Weinberg (Sili3, 15 et 57; $P > 0.05$).

Identification and characterization of 18 novel polymorphic microsatellite makers derived from expressed sequence tags in the Pacific oyster *Crassostrea gigas*.

C. SAUVAGE¹, P. BOUDRY^{1§}, S. LAPEGUE^{1*}

¹ Ifremer - Laboratoire de Génétique et Pathologie - La Tremblade, France

§ Present address : Ifremer –UMR M100 Physiologie et Ecophysiologie des Mollusques Marins – Plouzané, France.

* Correspondance :

Tél : +33 (0)5 46 76 26 30 ; Fax : +33 (0)5 46 76 26 11

email : slapegue@ifremer.fr

Keywords: EST-SSR markers, Null alleles, Pacific oyster, *Crassostrea gigas*

Abstract

We report the development of eighteen new polymorphic microsatellite DNA markers derived from *Crassostrea gigas* expressed sequence tags. Genotyping of 48 wild adult oysters sampled from Marennes-Oléron bay (France) revealed 12 to 48 alleles per locus. Observed and expected heterozygosity ranged from 0.64 to 1 and from 0.77 to 0.97 respectively. The development of these new markers creates a useful complementary tool for population genetics studies, parentage analysis and mapping in Pacific oyster, a species of major aquacultural and ecological importance.

The Pacific cupped oyster is one of the most widely cultivated aquatic species with a world annual production of 4.5 millions metric tons in 2005 (FAO, 2006). It has been introduced into many countries, and may naturalise and become invasive. Most production is based on collection of wild seed and this species has therefore benefited little from genetic improvement by selective breeding. A significant genetic basis has been shown for several traits of aquacultural interest, such as resistance to summer mortality (Degremont et al., 2007), indicating the potential for genetic improvement of cultured stocks. Moreover, the development of genomic resources is increasing (Hedgecock et al., 2005; Jenny et al., 2007; Tanguy et al., 2007). Current linkage maps include around 100 AFLP (Li et al., 2004) or 100 microsatellite markers (Hubert et al., 2004). Higher density maps, required for QTL or genome scan studies, rely on the development of additional microsatellite (Yu and Li, 2007; Wang et al., 2008) or SNP (Sauvage et al., 2007) markers. However, high frequencies of null alleles (Hedgecock et al., 2004) and segregation distortions (Launey and Hegecock, 2004) have been frequently observed in *C. gigas*, leading to additional constraints to the development of high density linkage maps.

Here we present the identification and the characterization of a new set of 18 EST-SSR markers, recently reported to present lower frequencies of null alleles than genomic SSRs (Yu and Li, 2008). A total number of 9272 EST contigs assembled from the Public sequences and the Marine Genomics database (<http://www.sigenae.org/>) were screened to identify every repeated tandem sequence (from penta to dinucleotide) using the Cotton Microsatellite Database Simple Sequence Repeat Server (Clemson University Genomics Institute; http://www.mainlab.clemson.edu/cm/ssr_server/). A total of 22 primers pairs were designed from the available sequences (see Table 1 for Accession numbers) using the Primer 3 program (Rozen, 2000; <http://frodo.wi.mit.edu/> - table 1).

PCR amplifications were performed as follows: an initial denaturation step of 2 min at 94°C, then the annealing temperature is decreased, from 65 to 55°C, in increments of 1°C for every subsequent set of 2 cycles, and followed by an extension step of 1 min at 72 °C. When

the lowest annealing temperature was reached, it was used for the 25 remaining cycles. The PCR mix was composed of 0.3U *Taq* polymérase (New England Biolabs), 10 mM provided buffer, 1 mM MgCl₂, 2 mM dNTP (Eurogentec), 10mM of each primer and 100 ng genomic DNA in a final reaction volume of 10 µL. Capillary electrophoresis was performed on a ABI 3130xl genetic analyser (Applied Biosystems) and alleles were scored using the software Genemapper version 4 (Applied Biosystems).

Eighteen SSR out of 22 gave PCR products of expected size and showed only two alleles. Descriptive statistics (table 1) on the genotypes from the 48 adult samples were obtained using Genetix 4.1 (Belkhir et al., 1996). The number of alleles per locus ranged from 12 to 48, with mean of 29.9. The expected and observed heterozygosity estimates, ranged from 0.77 to 0.97 and from 0.64 to 1, respectively. These results are similar to those previously published by Huvet et al. (2000). Testing for Hardy-Weinberg Equilibrium (HWE) was carried out using the Markov chain method (100 000 steps; 10 000 dememorisation) using ARLEQUIN Version 3.1 (<http://cmpg.unibe.ch/software/arlequin3/>). Three out of 18 loci (Cgsili3, 15 and 57) showed a significant deviance from equilibrium. Such deviations are unusually frequent in oysters and were recently observed in a set of EST-SSR developed by Yu and Li (2007) in the Pacific oyster. In most cases, such deviations are due to the high frequency of null alleles (Hedgecock et al., 2004). Moreover, genotypic linkage disequilibrium was detected between three markers (Cgsili3, 6 and 7; $P > 0.05$). The null allele frequency (r) was computed using the Cervus program (Marshall et al., 1998; http://www.fieldgenetics.com/pages/aboutCervus_Overview.jsp) and results show that this frequency is low ($r > 0.5$ for 17% of loci), which is as expected because the frequency of null alleles in EST-derived SSRs is usually lower than in genomic SSRs due to lower mutation at splicing sites (Goldstein and Schlötterer, 1999).

These novel microsatellite loci, developed by an *in silico* screening approach will be useful for further population genetics, parentage analysis and linkage mapping studies in the Pacific oyster.

Acknowledgements

This work was supported by the EU projects Marine Genomics, Aquafirst and the Bio-Informatics platform SIGENAE. We thank Helen McCombie for editing the English.

Chapitre III – Développement des marqueurs moléculaires et génotypage

Table 1. EST-SSR isolated in *Crassostrea gigas* with: primer sequences, number of individuals genotyped (Ni), observed heterozygosity (Ho), expected heterozygosity (He), testing of Hardy-Weinberg Equilibrium (HWE) and calculated frequency of null alleles in 48 individuals.

Locus	Repeat sequence	Primers sequences (5'-3')	Ni	Na (size range, bp)	Ho	He	HWE	Null alleles Frequency (r)	Accession no.
Cgsili3	(TA) ₂₇	F: TGAAATTAACAACCCTTGGT R: TTGAATCTAAAGAATATGGCAA	39	36 (324-403)	0.641	0,9589	0.032*	+0.1995	AM857068
Cgsili4	(AG) ₂₆	F: GGTGCAGTAGTTGGAAACAT R: TCACATTTAACTAGCGCTCTC	43	45 (267-397)	0.7907	0,9594	0.431	+0.0976	AM854894
Cgsili6	(GA) ₂₆	F: ATGAACGTCCAAGTTCAGAC R: ACACATTTCTTATAAAGCC	46	48 (238-410)	0.913	0.97	0.221	+0.0301	AM854296
Cgsili7	(GA) ₂₄	F: ACAACGCTATCAGAACCATT R: ATCTCCCGGCAAGTATATG	44	32 (209-300)	0.8864	0.9122	0.080	+0.0077	CU682571
Cgsili12	(CAATG) ₅	F: CCATCTTCTGTTTGGTTCAT R: GCACCGGATGGTAGATATT	45	40 (155-242)	1.0000	0.9644	0.057	-0.0183	AM857087
Cgsili15	(TCCA) ₆	F: TCTCTTCCATCCATCAATC R: GGTGATATTGTAACGGCATT	42	16 (200-230)	0.595	0.763	0.021*	+0.1289	AM856504
Cgsili26	(GA) ₁₃ GC(GA) ₇	F: CCCACTAAAGCTACCTATTGA R: TGTGGATTCCCTTAACTGACC	47	30 (202-278)	1.0000	0.9124	0.689	-0.0522	AM855372
Cgsili29	(AG) ₂₀	F: TTTATATGCGATTTCTGAAGC R: TTGACTAAGAATTCATCGGG	42	38 (302-400)	0.9524	0.9592	0.744	+0.0030	AM856933
Cgsili37	(TC) ₁₅	F: TTGCTGGTTGTGATGAATAG R: ATATCTGGCCTAACATGTGC	37	32 (109-190)	1.0000	0.9592	0.451	-0.0210	BQ427164
Cgsili38	(TG) ₆ C(GT) ₈	F: TTTCTGCTCTGTGTGTGTGT R: ACGTGGAGTGACGCTACTAT	34	23 (306-391)	0.9412	0.9325	0.332	-0.0060	AM854072
Cgsili39	(AG) ₁₃	F: GACCATACAGCTCTGTCCAT R: GCTACTGAATGAGAATGGCT	29	30 (331-445)	1.0000	0.9566	0.094	-0.0237	AM854746
Cgsili43	(GA) ₁₀	F: AAATGCTGCAGAAATAATCC R: AGATGGCTACAGTGAAATGG	43	15 (340-358)	0.8837	0.8807	0.077	-0.0013	AM854072
Cgsili44	(AG) ₇ AAA(GA) ₄	F: TGGCATTTCATGGTTAATTT R: TGTTGTATGAAATGTCGGAA	39	25 (208-261)	0.9744	0.9395	0.123	-0.0193	AM858556
Cgsili45	(AG) ₇ AAA(GA) ₄	F: GTCGAGAAATAAAGCTGGAA R: CAAGATTCCAAGGAAACAAA	41	20 (184-211)	1.0000	0.898	0.089	-0.0592	AM857706
Cgsili46	(TG) ₅ A(GT) ₇	F: CATGACAATCGAGTCCATAA R: CATGGTGGAGAAAGAGTTGT	39	33 (340-414)	0.9487	0.9481	0.091	-0.0010	AM856490
Cgsili50	(CA) ₁₀	F: CTATCTGAGCACGCTTCTCT R: TCTCTGTGATGATCTCAGG	46	25 (221-250)	0.848	0.913	0.078	+0.0314	AM865904
Cgsili56	(GTT) ₈	F: GTCCCTGTGTTACACTCCTG R: TATCAGTCCAACAAAGGAGG	42	21 (232-289)	1.0000	0.8747	0.239	-0.0720	CU682998
Cgsili57	(AAG) ₈	F: CAGTCCCTCTACGCTACATC R: ACACTACCGCTTTCCTGATA	35	12 (320-374)	1.0000	0.7727	0.046*	-0.1613	AM856592

* Significance level of 0.05

References

- Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N. and Bonhomme, F.: GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France), 1996.
- Degremont, L., Ernande, B., Bedier, E. and Boudry, P., 2007. Summer mortality of hatchery-produced Pacific oyster spat (*Crassostrea gigas*). I. Estimation of genetic parameters for survival and growth. *Aquaculture*. 262, 41-53.
- FAO (2006) Yearbook of Fisheries statistics: Summary Tables. World aquaculture production of fish, crustaceans, molluscs, etc by principal species in 2005. <http://www.fao.org/fi/statist/summtab/default.asp>.
- Goldstein, D. and Schlotterer, C.: *Microsatellites: Evolution and Applications*. Oxford University Press, 1999.
- Hedgecock, D., Gaffney, P.M., Gouletquer, P., Guo, X., Reece, K. and Warr Gregory, W., 2005. The case for Sequencing the Pacific Oyster Genome. *Journal of Shellfish Research*. 24, 429-441.
- Hedgecock, D., Li, G., Hubert S., Bucklin K., Ribes V., 2004. Widespread null alleles and poor cross-species amplification of microsatellite DNA cloned from the Pacific oyster *Crassostrea gigas*. *Journal of Shellfish Research* 23: 379-385.
- Hubert, S. and Hedgecock, D., 2004. Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. *Genetics*. 168, 351-62.
- Jenny, M.J., Chapman, R.W., Mancina, A., Chen, Y.A., McKillen, D.J., Trent, H., Lang, P., Escoubas, J.M., Bachere, E., Boulo, V., Liu, Z.J., Gross, P.S., Cunningham, C., Cupit, P.M., Tanguy, A., Guo, X., Moraga, D., Boutet, I., Huvet, A., De Guise, S., Almeida, J.S. and Warr, G.W., 2007. A cDNA microarray for *Crassostrea virginica* and *C. gigas*. *Marine Biotechnology* (NY). 9, 577-91.
- Launey, S. and Hedgecock, D., 2001. High genetic load in the Pacific oyster *Crassostrea gigas*. *Genetics*. 159, 255-65.
- Li, L. and Guo, X., 2004. AFLP-based genetic linkage maps of the Pacific oyster *Crassostrea gigas* (Thunberg, 1793). *Marine Biotechnology*. 6, 26-36.
- Marshall, T. C., J. Slate, L.E.B. Kruuk and J. M. Pemberton, 1998. "Statistical confidence for likelihood-based paternity inference in natural populations." *Molecular Ecology* 7(5): 639-655.
- Rozen, S. and H. Skaletsky, 2000. "Primer3 on the WWW for general users and for biologist programmers." *Methods Molecular Biology* 132: 365-86.
- Sauvage, C., Bierne, N., Lapegue, S. and Boudry, P., 2007. Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. *Gene*. 406, 13-22.
- Tanguy, A., Bierne, N., Saavedra, C., Pina, B., Bachere, E., Kube, M., Bazin, E., Bonhomme, F., Boudry, P., Boulo, V., Boutet, I., Cancela, L., Dossat, C., Favrel, P., Huvet, A., Jarque, S., Jollivet, D., Klages, S., Lapegue, S., Leite, R., Moal, J., Moraga, D., Reinhardt, R., Samain, J.-F., Zouros, E. and Canario, A., 2008. Increasing genomic information in bivalves through new EST collections in four species: Development of new genetic markers for environmental studies and genome evolution. *Gene*. 408, 27-36.
- Yu, H. and Li, Q.I., 2007. EST-SSR markers from the Pacific oyster, *Crassostrea gigas*. *Molecular Ecology Notes*. 7, 860-862.
- Yu, H. and Li, Q.I., 2008. Exploiting EST Databases for the Development and Characterization of EST-SSRs in the Pacific Oyster (*Crassostrea gigas*). *Journal of Heredity*. 99: 208-214.
- Wang, Y., Ren, R. and Yu Z., 2008. Bioinformatic mining of EST-SSR loci in the Pacific oyster, *Crassostrea gigas*. *Animal Genetics*. 39: 287-289.

3.7.3 Développement de marqueurs moléculaires de type SNP

L'ensemble des résultats du développement de marqueurs moléculaires de type SNP ainsi que l'analyse du biais d'usage du code génétique (ou utilisation de préférentielle de codons) chez l'huître creuse *C. gigas* sont détaillés au sein d'un article présenté dans les pages suivantes (Sauvage et al., 2007). L'encadré 2 résume les résultats obtenus pour le développement de SNP (figures 25 et 26).

Encadré 2: Synthèse des résultats du développement des marqueurs SNP

Les travaux de recherche de SNP ont permis d'obtenir les résultats suivants:

- Un total de 149 séquences d'EST ont été retenues pour identifier des sites polymorphes.
- Ces ESTs sont impliqués dans les mécanismes énergétiques et immunitaires chez l'huître creuse.
- Le taux de réussite de l'amplification par PCR et de séquençage a été limité de 41% (61/149) par la présence d'introns de grande taille, de produits de PCR non spécifique ou des difficultés techniques.
- 321 sites polymorphes ont été identifiés (Figure 25) :
 - 73,5% se situent dans les régions introniques et 26,5% dans les régions exoniques
 - dans les régions exoniques, 15,8% sont des mutations non-synonymes et 10,7% sont des mutations dites silencieuses ou synonymes

Figure 25 : Répartition et nature des marqueurs SNP identifiés.

- Le taux de polymorphisme nucléotidique a été estimé à :
 - un SNP toutes les 61 paires de bases dans les régions exoniques
 - un SNP toutes les 41 paires de bases dans les régions introniques
 - un ratio global ts/tv estimé à 1,3 (1,34 dans les régions codantes et à 1,17 dans les régions non codantes - Figure 26).

Figure 26 : Pourcentage de transitions et de transversions dans les régions codantes et non-codantes de l'ADN.



Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*

C. Sauvage^a, N. Bieme^b, S. Lapègue^{a,*}, P. Boudry^a

^a Laboratoire de Génétique et Pathologie - IFREMER - La Tremblade, France

^b Département de Biologie Intégrative, Institut des Sciences de l'Évolution de Montpellier UMR 5554 CNRS-UMI, Station Méditerranéenne de l'Environnement Littoral, Sète, France

Received 6 February 2007; received in revised form 14 May 2007; accepted 18 May 2007

Available online 2 June 2007

Abstract

DNA sequence polymorphism and codon usage bias were investigated in a set of 41 nuclear loci in the Pacific oyster *Crassostrea gigas*. Our results revealed a very high level of DNA polymorphism in oysters, in the order of magnitude of the highest levels reported in animals to date. A total of 290 single nucleotide polymorphisms (SNPs) were detected, 76 of which being localised in exons and 214 in non-coding regions. Average density of SNPs was estimated to be one SNP every 60 bp in coding regions and one every 40 bp in non-coding regions. Non-synonymous substitutions contributed substantially to the polymorphism observed in coding regions. The non-synonymous to silent diversity ratio was 0.16 on average, which is fairly higher to the ratio reported in other invertebrate species recognised to display large population sizes. Therefore, purifying selection does not appear to be as strong as it could have been expected for a species with a large effective population size. The level of non-synonymous diversity varied greatly from one gene to another, in accordance with varying selective constraints. We examined codon usage bias and its relationship with DNA polymorphism. The table of optimal codons was deduced from the analysis of an EST dataset, using EST counts as a rough assessment of gene expression. As recently observed in some other taxa, we found a strong and significant negative relationship between codon bias and non-synonymous diversity suggesting correlated selective constraints on synonymous and non-synonymous substitutions. Codon bias as measured by the frequency of optimal codons for expression might therefore provide a useful indicator of the level of constraint upon proteins in the oyster genome.

© 2007 Elsevier B.V. All rights reserved.

Keywords: SNP; Genetic diversity; Codon bias; *Crassostrea gigas*

1. Introduction

Single Nucleotide Polymorphisms (SNPs) are the most abundant sequence variations encountered in a genome (Cho et al., 1999; Picoult-Newberg et al., 1999; Griffin and Smith, 2000). Although moderately sparse in the human genome with one SNP per kb (Sachidanandam et al., 2001), they sometimes

reach high densities in some highly diverse species such as in some insects (e.g. one SNP every 125 bp in the *Aedes* mosquito genome, Morlais and Severson, 2003) or some crops (e.g. one SNP every 104 bp in the maize genome, Tenaillon et al., 2001). With the development of DNA-based marker assays and high-throughput genotyping technologies, SNPs have become markers of choice for large scale mapping and genotyping (Rafalski, 2002; Black et al., 2001). For example, it has had a great impact on the generation of genetic maps, for the analysis of genetic diversity, trait mapping and diagnostics. They are especially useful for association studies because of their high frequency in the genome, and they are genetically more stable than microsatellite markers. SNPs are therefore ideally suited for the generation of high-density genetic maps (Cho et al., 1999; Nairz et al., 2002) and have number of advantages for

Abbreviations: AFLP, Amplified Fragment Length Polymorphism; bp, base pair; kb, kilo base pair; Fop, Frequency of optimal codon; GC3, Percentage of G and C content at third position of codon; ORF, Open reading Frames; QTL, Quantitative Trait Loci; SNP, Single Nucleotide Polymorphism; SSR, Single Strand Repeat; ts, transition; tv, transversion.

* Corresponding author. LGP, Station IFREMER, 17390 La Tremblade, France. Tél.: +33 5 46 76 26 30; fax: +33 5 46 76 26 11.

E-mail address: slapegue@ifremer.fr (S. Lapègue).

0378-1119/\$ - see front matter © 2007 Elsevier B.V. All rights reserved.

doi:10.1016/j.gene.2007.05.011

population genetics studies (Vignal et al., 2002). However, only few SNP markers have been developed in marine bivalves and more generally in marine invertebrates.

Marine molluscs are recognised to reveal one of the highest level of allozyme polymorphism within the animal kingdom (average heterozygosity 15–30%, Ward et al., 1992; Bazin et al., 2006). The extreme heterozygosity of marine molluscs is at first best explained by large effective population sizes expected for species with high fecundities, extensive larval dispersal, dense populations and broad distribution. However, numerous studies have questioned the neutrality of allozyme variation in marine molluscs. Popular examples of direct selection on some allozyme loci come from the marine bivalves literature (Koehn et al., 1980; Karl and Avise, 1992; Riginos et al., 2002), although they have sometimes been criticized (McDonald, 1996; Bieme et al., 2003). The observation that multilocus heterozygosity is frequently correlated with fitness-related traits in these species (review in David, 1998) was initially taken as evidence for overdominance at allozymes (Mitton, 1993). However, a consensus now emerged that the correlation is in fact best explained by the indirect effect of deleterious mutations on neutral marker variations (David, 1998). This neutral alternative introduced two requisites that have a bearing on our understanding of the population genetics of marine bivalves: (i) a high genetic load and (ii) the existence of particular population structures generating a strong variance in individual inbreeding (Bieme et al., 2000). Both have received support. The genetic load has been quantified and was estimated to be extremely high (Bieme et al., 1998; Launey and Hedgecock, 2001). Several authors have challenged the idea that marine bivalves occur in large, homogeneous, randomly mating populations. Hedgecock's sweepstakes hypothesis proposes that large variation in reproductive success lead to effective population sizes several orders of magnitude below census numbers (Hedgecock, 1994). The slight, unpredictable but significant genetic differentiation observed at small spatial scales (Johnson and Black, 1984; David et al., 1997) is also consistent with the fragmentation of marine populations into small, transient reproductive groups. Finally, Bazin et al. (2006) reported that mitochondrial diversity was not higher in marine molluscs than in other taxonomic groups in sharp contrasts with allozyme diversity. Bazin et al. (2006) also reported a positive correlation between allozyme diversity and nuclear DNA diversity at a large taxonomic scale, suggesting that these two types of variation share the same information, namely population size. However, their meta-analysis nonetheless revealed that quantification of nuclear DNA diversity remains uncommon in many invertebrate groups, including molluscs, precluding the use of DNA polymorphism for comparison with mitochondrial diversity. Taking into consideration the doubts accumulated on the neutral status of allozymes and on the effective population size in marine bivalves, one might enquire thorough assessments of silent DNA polymorphism in these taxa.

Here, we describe the characterization of SNPs in coding and non-coding sequences of the Pacific oyster *Crassostrea gigas*. We used primer sequences designed in a set of 41 ESTs and direct sequencing of PCR products. The validity of detected

SNPs was ascertained by multiple sampling of the same allele in a sample of related individuals. This sampling strategy allowed us to remove technical artefacts of PCR and sequencing that generate singleton substitutions. We describe levels of diversity observed in coding and non-coding regions. We also undertake an investigation of codon usage bias and its relationship with DNA polymorphism. It is now widely accepted that selection acts on synonymous codons to improve translation in many organisms (Ikemura, 1992; Akashi, 1994) but not always (Duret, 2002). Variation in the effectiveness of selection on synonymous codons between species is likely to reflect variation in effective population sizes rather than variation in coefficients of selection (Li, 1987; Bulmer, 1991; Akashi, 1995; Cutter et al., 2006). Indeed, genomes with the strongest variation in codon usage correspond to species recognised to have large effective population sizes such as bacteria, yeast or insects (Ikemura, 1982; Merkl, 2003; Akashi, 1995) while natural selection does not appear to play a role in Mammals (Urrutia and Hurst, 2001). Although selection on synonymous codon use has been known for more than two decades, it remains unclear whether codon usage primarily affects the elongation rate or the fidelity of protein synthesis (Akashi, 2001; Duret, 2002). However, the latter hypothesis has received evidences in *Drosophila melanogaster*, *Caenorhabditis elegans* and *Escherichia coli* (Akashi, 1994; Marais and Duret, 2001; Stoletzki and Eyre-Walker, 2007). A negative correlation between the rate of non-synonymous substitution and codon bias have been described in a number of species (Stoletzki and Eyre-Walker, 2007; Pal et al., 2001; Betancourt and Presgraves, 2002) and seems best explained by the effect of selection on the accuracy of translation (Bieme and Eyre-Walker, 2006; Stoletzki and Eyre-Walker, 2007). As a consequence, some authors have recently proposed that codon bias might sometimes be used as an indicator of the selective constraint acting on a protein (Stoletzki and Eyre-Walker, 2007; Plotkin et al., 2006) as others proposed to use codon bias as a measure of the level of gene expression (Sharp and Li, 1987; Karlin and Mrazek, 2000; Coghlan and Wolfe, 2000). We here provide results suggesting that codon bias could provide useful information about the level of constraint upon a gene in the oyster genome.

2. Materials and methods

2.1. Biological material and DNA extraction

Twenty four two year-old oysters (*C. gigas*) were used in this SNP discovery experiment. They were produced during the French MOREST program that studied the genetic basis of the summer mortality phenomenon as described in Dégremont et al. (2007). These 24 individuals belongs to the third stage of selection for the resistance to the summer mortality and were selected to initiate a genetic linkage mapping experiment to detect quantitative trait loci (QTL) of aquacultural interest. These oyster are related with each others as following (Fig. 1): there are two replicates of four groups composed of three individuals (one male and two females), which are brothers and

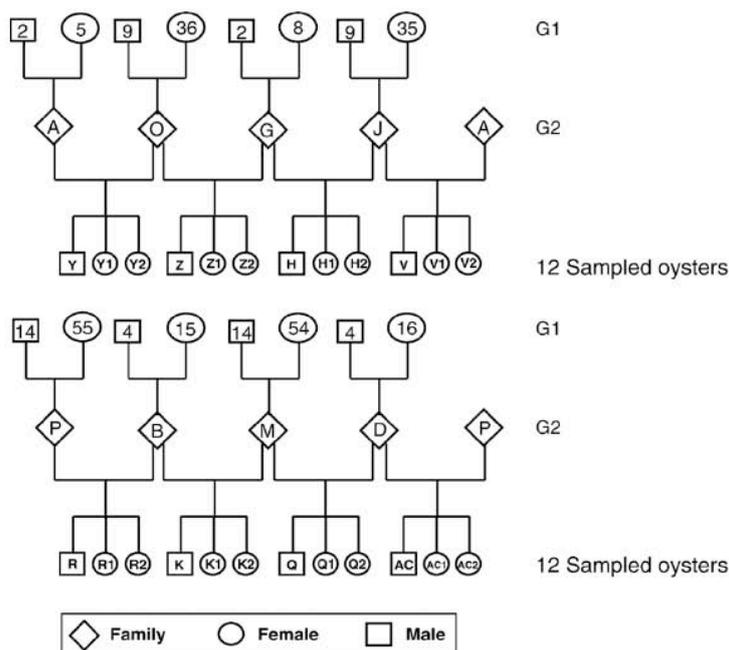


Fig. 1. Genealogical relationship between the oysters used in our analysis of polymorphism.

sisters. They are related by a semi-brother relationship to individuals of the three other groups. All individuals that compose the eight groups are the progeny issued from a cross of two individuals randomly taken from eight pools (G2) of individuals, named A, O, G, J, K, Q, R and AC. These eight pools are themselves the progeny of an ancestral couple of oysters (G1) selected for their resistance to summer mortality phenomenon (Dégremont et al., 2007).

DNA was extracted from samples of mantle tissue following the Promega DNA Wizard cleanup Kit recommendations. The quality of DNA was first checked on a 1.5% agarose gel (15 V/cm; 40 min) and secondly by quantification using an Eppendorf biospectrophotometer. This allowed to get 24 equal concentration of DNA (100 µg/mL) prepared for the amplification step.

2.2. Primer design and amplification

Primers were designed using the Primer3 software (Rozen and Skaletsky, 2000) from Expressed Sequence Tags (EST) developed in the Pacific oyster and retrieved from the Genbank database (<http://www.ncbi.nlm.nih.gov/>). Few additional unpublished ESTs were also added in our experiment (Huvet, Boulo and Renault, Pers. Com.). Primers produced fragments from 250 to 550 bp. PCR conditions were standardized by the use of a touchdown PCR protocol. The same reaction mix was used for each pair of primers and amplification was performed with three ranges of annealing temperatures (60 to 50 °C, 65 to 55 °C and 82 to 74 °C). The mix was composed for each sample of 0.3U *Taq* polymerase (New England Biolabs), 10 mM of provided buffer, 1 mM MgCl₂, 2 mM of dNTP (Eurogentec),

10 µM of each primer and 100 ng of genomic DNA, in a final reaction volume of 48 µL. Pairs of primers with the same optimal range of annealing temperature were grouped and ran on a Perkin-Elmer ABI 2700 PCR machine (Applied Biosystems) as following: initial denaturation step for 5 min at 94 °C, then 2 cycles of 1 min denaturation at 94 °C, for every subsequent 2 cycles, the annealing temperature was decreased by 1 °C during 1 min, and an extension step at 72 °C of 1 min. At the lowest annealing temperature (50, 55 or 74 °C), 25 cycles at 94 °C for 30 s, 30 s annealing and 1 min extension at 72 °C were applied. A quantity of 5 µL of PCR products was purified with 2 µL of the ExoSAP-IT enzyme (Amersham Biosciences) to remove non-incorporated dNTPs and primers according to the manufacturer manual. Then, purified PCR products were sequenced in a single direction using the forward primer with the ABI Prism BigDye v3 Terminator Cycle sequencing Kit (Applied Biosystems) and the sequences were analysed on an ABI 3100 *Avant* genetic analyser (Applied Biosystems).

2.3. Nucleotide diversity

Sequences were edited, corrected by hand if needed and aligned using Seqscape v2.1 software (Applied Biosystems) using the KB basecaller algorithm. After a Blast homology search (<http://www.ncbi.nlm.nih.gov/BLAST/>), the sequences that did not correspond, at least partly to the EST used to design primers, were removed. Multiple sequences alignment was exported to BioEdit v7.0.5. (Hall, 1997) and aligned with the corresponding EST sequence in order to infer intronic regions. Open Reading Frames (ORF) were inferred by trying the three frames and

Table 1
Nucleotide polymorphism in *Crassostrea gigas* nuclear genes

Genes	Accession number	Pn	Ps	Pnc	Type of non-coding sequence	Ln (bp)	Ls (bp)	Lnc (bp)	Lclds	π_s	π_n	π_{nc}	π_{si}	π_n/π_{si}	Fop	GC3
Drac 3	EF694083	1	1	2	Intron	122	37	78	325	0.027	0.008	0.026	0.026	0.314	0.39	0.71
Elongation factor 1alpha	EF694084	0	1	5	Intron	73	20	99	1386	0.050	0	0.051	0.050	0	0.42	0.48
Adrenal gland protein	EF694077	0	4	5	Intron	114	30	220	714	0.133	0.000	0.023	0.036	0.000	0.33	0.52
Astacin	EF694085	2	1	16	Intron	91	26	175	752	0.038	0.022	0.091	0.085	0.260	0.26	0.39
Cytochrome P450	EF694086	0	1	7	Intron	63	21	115	151	0.048	0	0.061	0.059	0	0.35	0.6
Tubulin	EF694087	2	0	4	Intron	78	27	142	533	0	0.026	0.028	0.024	1.083	0.28	0.46
GlycoProtein Hormone receptor	EF694088	1	1	0	UTR	207	56	67	3279	0.018	0.005	0.000	0.008	0.594	0.3	0.49
Glutaryl-CoA dehydrogenase-like protein	EF694089	0	0	18	UTR	10	32	202	216	0	0	0.089	0.077	0	0.3	0.41
Glutathion S transferase	EF694090	0	0	11	Intron	70	23	277	537	0	0	0.040	0.037	0	0.37	0.53
Amylase gene B	EF682217	1	1	0	Intron	224	73	88	1557	0.014	0.004	0.000	0.006	0.719	0.34	0.49
Apolipoprotein	EF694091	0	0	0	Intron	265	74	88	562	0	0	0.000	0		0.31	0.39
Ciao1 like protein	EF694092	2	6	6	Intron	154	40	107	564	0.150	0.013	0.056	0.082	0.159	0.27	0.38
Vasa-like	EF694093	0	1	9	Intron	138	48	303	2274	0.021	0	0.030	0.028	0	0.35	0.43
ATPase H ⁺	EF694094	0	0	6	Intron	61	17	141	183	0	0	0.043	0.038	0	0.37	0.64
BQ427367	EF694095	0	3	15	Intron	79	23	187	489	0.130	0	0.080	0.086	0	0.36	0.36
S-adenosylhomocysteine hydrolase	EF694096	0	0	12	Intron	97	32	145	470	0	0	0.083	0.068	0	0.28	0.52
Super oxyde dismutase	EF694097	0	1	2	Intron	97	32	145	576	0.031	0	0.014	0.017	0	0.41	0.52
Glucose 6 phosphatase	EF694098	0	1	5	Intron	100	29	117	130	0.034	0	0.043	0.041	0	0.51	0.58
Ik cytokine — down regulator of HLA	EF999946	0	0	11	Intron	44	7	250	312	0	0	0.044	0.043	0	0.31	0.59
Similar to sodium/potassium transporting ATPase	EF627977	0	0	0	Intron	79	20	98	100	0	0	0.000	0		0.41	0.52
Phospholipase C	EF999947	0	0	2	Intron	76	23	201	343	0	0	0.010	0.009	0	0.32	0.55
Calcium dependant protein kinase	EF694100	0	0	5	Intron	131	34	177	744	0	0	0.028	0.024	0	0.31	0.51
Inhibitor protein kappa B	EF694101	0	0	0	Intron	42	15	166	1086	0	0	0.000	0		0.27	0.41
Plastin	AF075690	1	0			96	27		177	0	0.010	0		0.19	0.36	
Actophorin	AF075694	0	0			81	21		201	0	0	0		0.25	0.35	
Myosin light chain	AJ563458	2	0			95	28		477	0	0.021	0		0.31	0.54	
Amylase gene A	EF694074	4	6			206	67		1560	0.090	0.019	0.090	0.217	0.32	0.51	
Beta tubuline	AY713400	0	0			118	38		1128	0	0	0		0.46	0.54	
BQ426586	BQ426586	2	1			132	39		279	0.026	0.015	0.026	0.591	0.29	0.49	
Laccase	EF999948	0	4			92	28		321	0.143	0	0.143	0	0.33	0.5	
Similar to macrophage expressed gene 1	EF627979	1	3			131	37		471	0.081	0.008	0.081	0.094	0.32	0.52	
Notch	EF999949	1	0			159	42		310	0	0.006	0		0.27	0.37	
Flavin-containing monooxygenase 2	AJ585074	0	2			130	35		1365	0.057	0	0.057	0	0.39	0.61	
Integrin	EF694075			30	UTR			241	234			0.124	0.124	0.37	0.53	
Gigasins 2 (exons 1-2)	EF694076			0	Intron			162	83			0.000	0	0.12	0.21	
Paired type homeodomain protein	AY187692			11	Intron			168	96			0.065	0.065	0.39	0.67	
Cyclophilin	EF694080			8	Intron			199	492			0.040	0.040	0.37	0.5	
Scavenger	EF694081			2	Intron			168	246			0.012	0.012	0.14	0.29	
HA114	EF694082			5	Intron			256	360			0.020	0.020	0.38	0.53	
Kazal-type proteinase inhibitor	CB617337			25	UTR			404	395			0.062	0.062	0.21	0.35	
Arha2	EF694078			3	Intron			261	579			0.011	0.011	0.42	0.6	

Pn : Number of non-synonymous polymorphic sites in coding DNA.

Ps : Number of synonymous polymorphic sites in coding DNA.

Pnc : Number of polymorphic sites in non-coding DNA.

Ln : Number of non-synonymous sites in coding DNA.

Ls : Number of synonymous sites in coding DNA.

Lclds : Length of the longest coding sequence available (bp).

Lnc : Length of the non-coding sequence (bp).

π_n : Number of non-synonymous SNPs per non-synonymous sites, Pn/Ln.

π_s : Number of synonymous SNPs per synonymous sites, Ps/Ls.

π_{nc} : Number of non-coding SNPs per non-coding sites, Pnc/Lnc.

π_{si} : Number of silent SNPs per silent sites, (Ps+Pnc)/(Ls+Lnc).

Fop: Frequency of optimal codon.

GC3: G/C content at third position of codon.

finding the longest ORF together with the help of orthologous sequence from GenBank. Regions outside the ORF but matching the EST sequence were considered to be UTRs. However, UTRs were found for only 4 loci (Table 1) and the majority of non-coding DNA sequences described in the present study belongs to

introns. We expected to find identical alleles by descent (i.e. the same sequence) several times because individuals of the sample are related. As a consequence, singleton substitutions were inferred to be technical artefacts arisen during PCR and sequencing steps. Although, SNPs were characterised by direct

sequencing of PCR products, this sampling strategy allowed us to obtain a valid characterisation of these markers. SNPs were identified as transitions or transversions for both coding and non-coding regions. For SNPs occurring in coding sequences, variations were classified as synonymous or non-synonymous changes. The analysis of genetic diversity was conducted with DnaSP 4.0 (Rozas and Rozas, 1995). We computed the average number of SNPs per site (Π) which is the number of polymorphic sites divided by the length of the sequence. Calculations were conducted independently for non-coding (Π_{nc}), synonymous (Π_s), and non synonymous (Π_{ns}) substitutions. In order to compute Π_s and Π_{ns} , we estimated the number of synonymous and non-synonymous sites in a coding sequence with the method of Nei and Gojoberi (1986).

2.4. Synonymous codon use

In order to investigate codon usage bias, the table of optimal codons was deduced from the analysis of an EST dataset using EST counts as a rough assessment of gene expression (Duret and Mouchiroud, 1999). The dataset was composed of 8800 EST sequences which were available for *C. gigas* in Genbank (<http://www.ncbi.nlm.nih.gov/>) and the Marine Genomics Europe Network of Excellence databases. Open reading frames (ORFs) were identified by choosing the longest possible translation into amino acid sequence. Sequences with an ORF smaller than 100 codons (300 bp) were removed from the dataset to prevent the occurrence of wrong ORFs and to have enough codons to compute the frequency of synonymous codons. It is likely that some EST annotations are wrong as a

result of artifactual frameshift substitutions in single-pass sequences. However, our procedure should result in removing a part of the coding region from the analysis rather than misinferring UTRs as coding regions. We therefore argue that sequence quality problems could introduce a statistical noise but should not bias the analysis. We built clusters of ESTs corresponding to the same gene by using the algorithm developed by Bazin et al. (2005) to construct the Polymorphix database. Two sequences were clustered into the same family if they shared fragment longer than 300 pb with similarity superior to 95%. Only the longest ORF of a cluster was used to compute tables of codon usage. The frequency of synonymous codon per amino acid was computed with the program CODONW (Peden, 1999). A synonymous codon was inferred as “optimal” when its frequency significantly increased with EST counts (Duret and Mouchiroud, 1999). This was tested by non-parametric Spearman’s correlation in JMP v5.0© (SAS Institute Inc.). Once optimal codons were inferred, we used the frequency of optimal codons (Fop; Ikemura, 1985) as a measure of codon bias.

3. Results

3.1. Types and distribution of polymorphism

A total of 84 EST sequences were chosen for primer design. Fifteen of these ESTs were chosen on the basis of their putative function in relation to summer mortality. Forty one ESTs (51%) amplified a clear genomic DNA fragment. For 31 of the 41 amplifying loci (72%), the size of the fragment obtained was

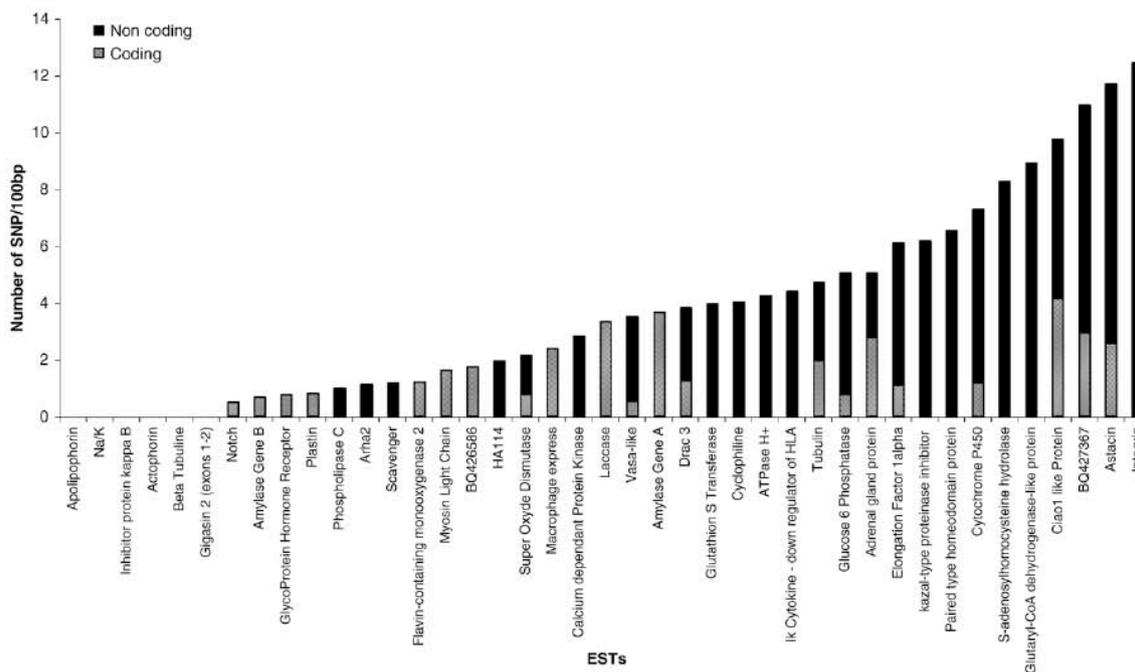


Fig. 2. Level of polymorphism detected in coding and non-coding parts of ESTs.

longer than the size expected from the EST sequence because of the presence of introns. Therefore, we suspect that the presence of several and/or very long introns could partly explain the rate of failure of PCR amplification.

Approximately, a total of 10.5 kb of the pacific oyster genome were sequenced with a total of 290 SNP markers identified in both coding and non-coding DNA. Around 3/4 of SNPs (69%) were localised in non-coding regions (Table 1). The level and the repartition of this polymorphism in each locus is shown in the Fig. 2. We found only 4 monomorphic locus which represented only 5% of amplified loci. The average density of SNPs in coding regions was estimated to 1 every 60 bp and 1 every 40 bp in non-coding ones. An average number of 6.7 SNPs were detected in each amplicon with a minimum of 0 (monomorphic sequence) and a maximum of 30 SNPs within a 241 bp (1 SNP every 9 bp) entirely non-coding sequence of the Integrin gene.

Our data show a global t_s/t_v ratio (1.3) in the Pacific oyster, which is similar to the ratio observed in *Drosophila* (1.5, Moriyama and Powell, 1996) or humans (1.4, Brookes, 1999). t_s/t_v was equal to 2.1 for synonymous changes while it was 0.9 and 1.2 for non-synonymous and non-coding changes respectively (exact t Fisher test, $t=0.264$). With twice more possibilities of transversions than transitions, t_s/t_v should ideally be equal to 1/2, if substitutions appeared randomly in DNA. However, it is well-known that a mutational bias favours transition over transversion. Because of the structure of the genetic code, synonymous changes are more often transitions than transversions. As much of the diversity is composed of synonymous substitutions, the t_s/t_v ratio increases in coding sequences ($t_s/t_v=1.34$). The same phenomenon is observed in the mosquito *Aedes aegypti* where t_s/t_v in coding sequences is similar. (Morlais and Severson, 2003). The mutational bias for

transition is therefore more accurately estimated from non-coding sequences and was estimated to be $t_s/t_v=1.2$ in *C. gigas*.

The number of SNPs per site was calculated for both coding and non-coding regions. Diversities were similar at synonymous ($\Pi_s=0.035$) and non-coding positions ($\Pi_{nc}=0.038$) and were not significantly different (per permutation t test $=-0.55$, $P=0.29$). Although the level of polymorphism in coding and non coding regions was positively correlated as expected for neutral substitutions sharing correlated genealogies, the correlation was not significant (Spearman's $\rho=0.3$; $P=0.16$). As the two kinds of silent substitutions gave the same information, they have been combined into a single indice (Π_{si}). The non-synonymous to silent polymorphism ratio (Π_n/Π_{si}) was 0.16 on average. However, Π_n/Π_{si} varied greatly from one gene to another.

3.2. Codon usage bias

Table 2 presents Spearman's correlation coefficients between the frequency of synonymous codons per amino acid and EST counts. As an illustration, we explain the results obtained at the Phenylalanine amino acid. Phenylalanine is a two-fold degenerate codon, which means that two different codons (UUU and UUC) can code for this amino acid. The average frequency of this amino acid among all others is 0.05 in the Pacific oyster genome and it does not vary significantly with expression levels as measured by EST counts. Ideally, each of the two codons of the phenylalanine should be equally represented by a frequency of 0.5. However, as expected in a GC-poor genome, the UUU codon is over-represented on average in lowly expressed genes (~60% in the category of genes with a single EST). On the other hand, the frequency of the UUC codon increases with expression level and the situation is reversed in highly expressed genes for

Table 2
Table of optimal codons deduced from the analysis of the EST dataset

Fourfold degenerate codons				Twofold degenerate codons				Six and threefold degenerate codons					
Aa	Codon	Spearman correlation	Probability	aa	Codon	Spearman correlation	Probability	aa	Codon	Spearman correlation	Probability		
Thr	ACU	-0.02	NS	Lys	AAA			Arg	AGA	-0.03	NS		
	ACC	0.04	0.012		AAG	0.06	<.001		AGG	0.03	NS		
	ACA	-0.01	NS		AAU				CGU	0.04	0.012		
	ACG	-0.02	NS		AAC	0.04	0.01		CGC	0.05	0.007		
Pro	CCU	0.01	NS	Gln	CAA			CGA	0.01	NS			
	CCC	0.04	0.023		CAG	0.07	<.001		CGG	-0.01	NS		
	CCA	-0.02	NS		His	CAU				Leu	CUU	-0.02	NS
	CCG	-0.03	NS			CAC	0.03		0.023		CUC	0.05	0.002
Ala	GCU	0.04	0.011	Glu		GAA			CUA		-0.03	NS	
	GCC	0.05	0.003			GAG	0.02	NS	CUG		0.06	0.001	
	GCA	-0.05	0.006		Asp	GAU			UUA	-0.05	0.003		
	GCG	-0.02	NS			GAC	0.03	NS	UUG	0.01	NS		
Gly	GGU	-0.02	NS	Tyr		UAU			Ser	AGU	-0.02	NS	
	GGC	0.01	NS			UAC	0.02	NS		AGC	0.01	NS	
	GGA	0.05	0.008		Cys	UGU				UCU	0	NS	
	GGG	-0.02	NS			UGC	0.01	NS		UCC	0.05	0.005	
Val	GUU	-0.03	NS	Phe		UUU			UCA	-0.01	NS		
	GUC	0.05	0.002			UUC	0.07	<.001	UCG	-0.03	NS		
	GUA	-0.05	0.002		Ile				AUU	0	NS		
	GUG	0.04	0.011			AUC	0.1	<.001	AUA	-0.1	<.001		

which the UUU codon is under-represented (~40% in the category of genes with 5 or more ESTs). The same trends were observed for other amino-acid: although -A and -U ending codons are more frequent on average as expected in a GC-poor genome, -G or -C ending optimal codons are more frequently observed (Table 2). As a consequence, GC-content at the third coding position is also correlated to expression levels (Spearman's $\rho=0.087$, $P<0.0001$) although less than *Fop* (Spearman's $\rho=0.13$, $P<0.0001$). As a consequence, codon usage is paradoxically less balanced in lowly expressed genes that are GC-poor than in highly expressed genes that reach an intermediate GC-content. This result highlights that codon usage bias is difficult to establish without an assessment of expression levels (Duret and Mouchiroud, 1999). In the case of the genome of *C. gigas*, a simple analysis of codon usage without expression data could have resulted in the inference that GC-poor genes are more biased and that -A and -U ending codons are optimal for translation while it is indeed the reverse – translational selection primarily favours -G or -C ending codons although rare in a GC-poor genome but the effectiveness of selection is not sufficient to result in a strong enrichment in GC-content which remains intermediate in biased genes. The results we obtained in *C. gigas* are similar to those recently reported in some GC-poor Nematode genomes in which translational selection nonetheless tends to favour -G or -C ending codons (Cutter et al., 2006).

We analysed whether genetic diversities would be correlated with codon bias. Synonymous diversity could have been expected to decrease with codon bias because of selection for codon usage. However, selection on synonymous substitutions is often too small to result in a detectable correlation (Bulmer, 1991; McVean and Charlesworth, 2000). In *Drosophila* for instance, although synonymous divergence decrease with codon bias (Sharp and Li, 1987) synonymous

polymorphism does not vary with codon bias with a dataset size comparable to this study (Bierne and Eyre-Walker, 2006). On the other hand, we detected a significant correlation between non-synonymous diversity and *Fop* (Fig. 3; Spearman's $\rho=-0.47$, $P=0.005$) or between Π_n/Π_{si} and *Fop* (Spearman's $\rho=-0.45$, $P=0.02$).

4. Discussion

4.1. High nucleotide diversity

The level of DNA polymorphism observed in this study is one of the highest ever observed to date, with one SNP every 40 bp in non-coding regions. To our knowledge, the highest levels of DNA polymorphism reported in the animal kingdom were found in the nematode *Caenorhabditis remanei* (Cutter et al., 2006) and the sea squirts *Ciona savignyi* (Small et al., 2007) with one SNP every 20 bp. Insect species are also recognised to often be highly polymorphic. The density of SNPs was reported to be one every 50 bp in *Drosophila* (Shapiro et al., 2007) and one every 125 bp in Mosquitoes (Morlais and Severson, 2003). Our results therefore confirm previous reports on allozyme diversities which revealed that marine invertebrates are amongst the most diverse animal species (Ward et al., 1992; Solé-Cava and Thorpe, 1991). Furthermore, DNA sequence polymorphism provides the opportunity to compare different categories of substitutions on which selection is expected to act differently. Synonymous and non-coding (silent) substitutions that do not result in a change in the protein can be used as a neutral reference to infer the level of purifying selection that act on non-synonymous (replacement) substitutions. A theoretical expectation is that selection should more efficiently remove deleterious mutations in species with large population sizes (Kimura, 1983). The non-synonymous to silent polymorphism ratio (Π_n/Π_{si}) should therefore be inversely correlated to the effective population size. In addition, unlike divergence data, polymorphism data should not be affected by adaptive evolution (McDonald and Kreitman, 1991) and Π_n/Π_{si} is expected to be a good measure of the proportion of non-synonymous substitutions that are effectively neutral (Smith and Eyre-Walker, 2002). We found that non-synonymous substitutions contributed substantially to the polymorphism observed in *C. gigas*: $\Pi_n/\Pi_{si}=16\%$. This value is much lower than in humans (0.22, Bustamante et al., 2005), but it is fairly higher to what has been reported in *Drosophila* (0.12, Shapiro et al., 2007) and much higher than in *Caenorhabditis remanei* (0.09, Cutter et al., 2006) and *Ciona savignyi* (0.07, Small et al., 2007). Therefore, purifying selection does not appear to be as strong as it could have been expected for a species with a large effective population size. However, this result at the molecular level concurs with the observation of a high genetic load in this species (Launey and Hedgecock, 2001) and would require further attention. Unfortunately, without any data on the frequency of substitutions in natural populations, it is difficult to investigate further the potentially deleterious nature of segregating non-synonymous polymorphisms (e.g. Eyre-Walker et al., 2006).

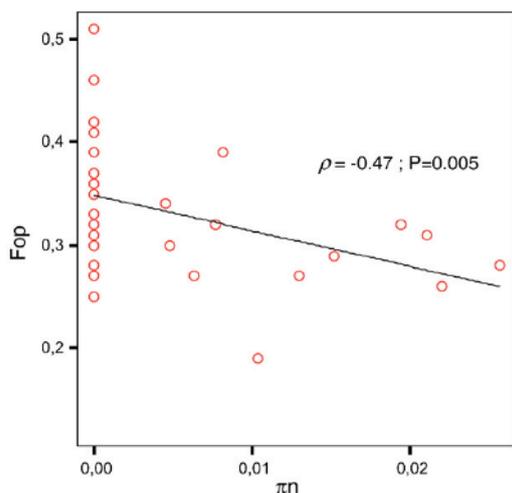


Fig. 3. Correlation between non synonymous diversity and codon bias revealed by *Fop* (Frequency of Optimal Codon).

Some polymorphisms that affect biological functions can occur outside of coding regions, in promoters or other regulating regions. Non-coding DNA contains important sequences for regulatory functions and can be constrained to some extent. Detectable levels of constraint on non-coding DNA have already been reported in *Drosophila* where non-coding diversity is lower than synonymous diversity (Moriyama and Powell, 1996; Andolfatto, 2005). The non-coding sequences we studied were in the close vicinity of genes, mostly in introns. However, contrary to what was reported in *Drosophila* we observed a similar level of diversity between the two kinds of silent substitutions. We conclude that synonymous and non-coding substitutions are equally (un)constrained in the oyster genome.

Recombination might be an interesting parameter to investigate further in the future in order to reconcile the possible discrepancy observed between high diversity and relaxed purifying selection. Indeed, intriguing results have been reported in marine bivalves. Guo and Allen (1996) reported the existence of a single recombinational hot spot in the proximal region of each chromosome arm in *Mulinia lateralis*. Hubert and Hedgecock (2004) detected chromosomal rearrangements in *C. gigas* and suggested that blocks of the genome may retain linkage disequilibria for longer than might be expected in putatively large, well-mixed populations.

4.2. Pattern of codon usage bias

The pattern of codon usage within a genome results from a balance among selection, random genetic drift and mutation (Bulmer, 1991). Our results in *Crassostrea gigas* first suggest that a mutational bias tends to enrich the genome in A and T because the genome is GC-poor on average (GC3=43%). Secondly, using a moderately large EST database (8800 ESTs) we nonetheless were able to easily detect synonymous codons that significantly increase with expression levels measured with EST counts (Duret and Mouchiroud, 1999). Ideally, one may verify that optimal codons correspond to the most abundant tRNAs in the cell (Ikemura, 1985). Unfortunately, tRNA concentrations in the cell or tRNA gene copy numbers are not available for *C. gigas* to date and we had to deal with expression levels alone. In addition, it should be noted here that expression levels inferred from EST data seem to be affected by gene length (Munoz et al., 2004). It can potentially be a problem because gene length and codon bias are correlated in a number of species, positively in Prokaryotes (Eyre-Walker, 1996) but negatively in Eukaryotes (Moriyama and Powell, 1998; Duret and Mouchiroud, 1999). Although puzzling the correlation is not necessarily the result of a direct link of causality (Duret and Mouchiroud, 1999). It seems difficult to imagine that a gene length effect could bias the inference of optimal codons for two reasons: (i) Longer genes are more energetically costly to translate. It seems expected that they could be underrepresented in the category of highly expressed genes. The fact that expression levels measured with EST data miss the correlation does not imply that other variables such as codon bias cannot be studied with EST

counts. (ii) Although a significant correlation between codon bias and gene length has been reported in some genomes, the correlation is always small, accounting for <5% of the total variance in codon bias (Moriyama and Powell, 1998).

Most of these codons that we inferred as being optimal for translation were -G or -C ending codons. As a consequence selection for translational efficiency tends to enrich highly expressed genes in G and C (GC3=49% for genes with five or more ESTs) when compared to lowly expressed genes (GC3=43% for genes with one EST). A difference of 6% in GC-content (or 5% in Fop) between highly and lowly expressed genes may be considered as small. However, EST count is a very rough assessment of expression level and the inherent statistical noise of an analysis of a moderately small EST dataset prevents from using this difference as a reliable estimate of selection intensity. Nonetheless, using a much larger EST database (114665 ESTs) in *Drosophila melanogaster* (data from Hey and Kliman, 2002), one may remark that genes with one EST are already GC-rich (GC3=64%) and that highly expressed genes are not tremendously richer in GC (GC3=70% for genes with more than 50 ESTs). Conversely, the simple fact that a significant effect is detected with a small EST dataset could have been taken as evidence that translational selection is strong in *C. gigas*.

We observed a strong and significant negative correlation between the frequency of optimal codons and non-synonymous diversity. Several hypotheses have been proposed to explain the correlation but an agreement recently emerged to propose that this is a consequence of correlated selective constraints on synonymous and non-synonymous substitutions (Bierne and Eyre-Walker, 2006; Stoletzki et al., 2005) possibly as a consequence of selection for translational robustness (Drummond et al., 2005). Indeed, Akashi (1994), Marais et al. (2004) and Stoletzki and Eyre-Walker (2007) provided evidence for selection on translational accuracy in *Drosophila*, *C. elegans* and *E. coli* respectively. Codon usage bias in *C. gigas* could therefore be due, at least partly, to selection for translational accuracy: a synonymous codon tends to be replaced by a more preferred one to reduce the fitness cost of a misincorporation (minimize the costs of proofreading and/or the cost of producing incorrect proteins). However, the confirmation of translational selection for accuracy in *C. gigas* would require further tests (see Stoletzki and Eyre-Walker, 2007). Whatever the exact cause of the correlation, our results nonetheless provide evidence that codon bias as measured by the frequency of optimal codons for expression might be used as a useful indicator of the selective constraint acting on a protein in the oyster genome (Stoletzki et al., 2005).

Acknowledgements

The authors are very indebted to Erick Bazin for the procedure allowing clustering EST sequences in gene families and thank two anonymous referees for their constructive remarks. This work was partly funded by the European project Aquafirst (Combined genetic and functional genomic approaches for stress and disease

resistance marker assisted selection in fish and shellfish, FP6), the Marine Genomics Europe Network of Excellence (FP6), the French national programme GIS “Génomique Marine” (The Bivalvomix Project, coord. N. Bierne), the Bureau des Ressources Génétiques (BRG) and the Region Poitou-Charentes (CPER). The authors would like to thank also Nicole Faury, Tristan Renault, Viviane Boulo and Amaud Huvet who provided us some of the primer sequences used in this paper.

References

- Akashi, H., 1994. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* 136, 927–935.
- Akashi, H., 1995. Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. *Genetics* 139, 1067–1076.
- Akashi, H., 2001. Gene expression and molecular evolution. *Curr. Opin. Genet. Dev.* 11, 660–666.
- Andolfatto, P., 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437, 1149–1152.
- Bazin, E., Duret, L., Penel, S., Galtier, N., 2005. Polymorphix: a sequence polymorphism database. *Nucleic Acids Res.* 33, 481–484.
- Bazin, E., Glemin, S., Galtier, N., 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science* 312, 570–572.
- Betancourt, A.J., Presgraves, D.C., 2002. Linkage limits the power of natural selection in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 13616–13620.
- Bierne, N., Eyre-Walker, A., 2006. Variation in synonymous codon use and DNA polymorphism within the *Drosophila* genome. *J. Evol. Biol.* 19, 1–11.
- Bierne, N., Launey, S., Naciri-Graven, Y., Bonhomme, F., 1998. Early effect of inbreeding as revealed by microsatellite analyses on *Ostrea edulis* larvae. *Genetics* 148, 1893–1906.
- Bierne, N., Tsitroni, A., David, P., 2000. An inbreeding model of associative overdominance during a population bottleneck. *Genetics* 155, 1981–1990.
- Bierne, N., Daguin, C., Bonhomme, F., David, P., Borsa, P., 2003. Direct selection on allozymes is not required to explain heterogeneity among marker loci across a *Mytilus* hybrid zone. *Mol. Ecol.* 12 (9), 2505–2510.
- Black, W.C.T., Baer, C.F., Antolin, M.F., DuTeau, N.M., 2001. Population genomics: genome-wide sampling of insect populations. *Annu. Rev. Entomol.* 46, 441–469.
- Brookes, A.J., 1999. The essence of SNPs. *Gene* 234, 177–186.
- Bulmer, M., 1991. The Selection-Mutation-Drift Theory of synonymous codon usage. *Genetics* 129, 897–907.
- Bustamante, C.D., et al., 2005. Natural selection on protein-coding genes in the human genome. *Nature* 437, 1153–1157.
- Cho, R.J., et al., 1999. Genome-wide mapping with biallelic markers in *Ara-bidopsis thaliana*. *Nat. Genet.* 23, 203–207.
- Coghlan, A., Wolfe, K.H., 2000. Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. *Yeast* 16, 1131–1145.
- Cutter, A.D., Baird, S.E., Charlesworth, D., 2006. High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. *Genetics* 174, 901–913.
- David, P., 1998. Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity* 80 (5), 531–537.
- David, P., Perdieu, M.A., Pernot, A.F., Jarne, P., 1997. Spatial and temporal population genetic structure in the marine bivalve *Spisula ovalis*. *Evolution* 51, 1318–1322.
- Dégremont, L., Ermande, B., Bedier, E., Boudry, P., 2007. Summer mortality of hatchery-produced Pacific oyster spat (*Crassostrea gigas*). I. Estimation of genetic parameters for survival and growth. *Aquaculture* 262, 41–53.
- Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O., Arnold, F.H., 2005. Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. U. S. A.* 102, 14338–14343.
- Duret, L., 2002. Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* 12, 640–649.
- Duret, L., Mouchiroud, D., 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4482–4487.
- Eyre-Walker, A., 1996. Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy? *Mol. Biol. Evol.* 13, 864–872.
- Eyre-Walker, A., Woolfit, M., Phelps, T., 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173, 891–900.
- Griffin, T.J., Smith, L.M., 2000. Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. *Trends Biotechnol.* 18, 77–84.
- Guo, X., Allen Jr., S.K., 1996. Complete interference and nonrandom distribution of meiotic crossover in a mollusc, *Mulinia lateralis* (Say). *Biol. Bull.* 191, 145–148.
- Hall, T., 1997. Bioedit. (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>).
- Hedgecock, D., 1994. Does variance in reproductive success limit effective population sizes of marine organisms? In: Beaumont, A.R. (Ed.), *Genetics and Evolution of Aquatic Organisms*. Chapman and Hall, London, pp. 122–134.
- Hey, J., Kliman, R.M., 2002. Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. *Genetics* 160, 595–608.
- Hubert, S., Hedgecock, D., 2004. Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. *Genetics* 168, 351–362.
- Ikemura, T., 1982. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. *J. Mol. Biol.* 158, 573–597.
- Ikemura, T., 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* 2, 13–34.
- Ikemura, T., 1992. In: Hatfield, D.L., Lee, B.J., Pirtle, R.M. (Eds.), *Correlation between codon usage and tRNA content in microorganisms, in Transfer RNA in Protein Synthesis*. CRC, Boca Raton, FL, pp. 87–111.
- Johnson, M.S., Black, R., 1984. Pattern beneath chaos: the effect of recruitment on genetic patchiness in an intertidal limpet. *Evolution* 38, 1371–1383.
- Karl, S.A., Avise, J.C., 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science* 256, 100–102.
- Karlin, S., Mrazek, J., 2000. Predicted highly expressed genes of diverse prokaryotic genomes. *J. Bacteriol.* 182, 5238–5250.
- Kimura, M., 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, New York.
- Koehn, R.K., Newell, R.L., Immermann, F., 1980. Maintenance of an aminopeptidase allele frequency cline by natural selection. *Proc. Natl. Acad. Sci. U. S. A.* 77, 5385–5389.
- Launey, S., Hedgecock, D., 2001. High genetic load in the Pacific oyster *Crassostrea gigas*. *Genetics* 159, 255–265.
- Li, W.H., 1987. Models of nearly neutral mutations with particular implications for non-random usage of synonymous codons. *J. Mol. Evol.* 24, 337–345.
- Marais, G., Duret, L., 2001. Synonymous codon usage, accuracy of translation, and gene length in *Caenorhabditis elegans*. *J. Mol. Evol.* 52, 275–280.
- Marais, G., Domazet-Lošo, T., Tautz, D., Charlesworth, B., 2004. Correlated evolution of synonymous and nonsynonymous sites in *Drosophila*. *J. Mol. Evol.* 59, 771–779.
- McDonald, J.H., 1996. Detecting non-neutral heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol. Biol. Evol.* 13, 253–260.
- McDonald, J.H., Kreitman, M., 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351, 652–654.
- McVean, G.A.T., Charlesworth, B., 2000. The effects of Hill–Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* 155, 929–944.
- Merk1, R., 2003. A survey of codon and amino acid frequency bias in microbial genomes focusing on translational efficiency. *J. Mol. Evol.* 57, 453–466.
- Mitton, J.B., 1993. Theory and data pertinent to the relationship between heterozygosity and fitness. In: Thornhill, N.W. (Ed.), *The Natural History of Inbreeding and Outbreeding—Theoretical and Empirical Perspectives*. University of Chicago Press, Chicago, pp. 17–41.
- Moriyama, E.N., Powell, J.R., 1996. Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* 13, 261–277.

- Moriyama, E.N., Powell, J.R., 1998. Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic Acids Res.* 26, 3188–3193.
- Morlais, I., Severson, D.W., 2003. Intraspecific DNA variation in nuclear genes of the mosquito *Aedes aegypti*. *Insect Mol. Biol.* 12, 631–639.
- Munoz, E.T., Bogarad, L.D., Deem, M.W., 2004. Microarray and EST database estimates of mRNA expression levels differ: the protein length versus expression curve for *C. elegans*. *BMC Genomics* 5, 30.
- Nairz, K., Stocker, H., Schindelholz, B., Hafen, E., 2002. High-resolution SNP mapping by denaturing HPLC. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10575–10580.
- Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426.
- Pal, C., Papp, B., Hurst, L.D., 2001. Highly expressed genes in yeast evolve slowly. *Genetics* 158, 927–931.
- Peden, J.F., 1999. *CodonW: Analysis of Codon Usage*. University of Nottingham.
- Picoult-Newberg, L., et al., 1999. Mining SNPs from EST databases. *Genome Res.* 9, 167–174.
- Plotkin, J.B., Dushoff, J., Desai, M.M., Fraser, H.B., 2006. Estimating selection pressures from limited comparative data. *Mol. Biol. Evol.* 23, 1457–1459.
- Rafalski, A., 2002. Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* 5, 94–100.
- Riginos, C., Sukhdeo, K., Cunningham, C.W., 2002. Evidence for selection at multiple allozyme loci across a mussel hybrid zone. *Mol. Biol. Evol.* 19, 347–351.
- Rozas, J., Rozas, R., 1995. DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Comput. Appl. Biosci.* 11, 621–625.
- Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132, 365–386.
- Sachidanandam, R., et al., 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928–933.
- Shapiro, J.A., et al., 2007. Adaptive genic evolution in the *Drosophila* genomes. *Proc. Natl. Acad. Sci. U. S. A.* 104, 2271–2276.
- Sharp, P.M., Li, W.H., 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. *Mol. Biol. Evol.* 4, 222–230.
- Small, K.S., Brudno, M., Hill, M.M., Sidow, A., 2007. Extreme genomic variation in a natural population. *Proc. Natl. Acad. Sci. U. S. A.* 104, 5698–5703.
- Smith, N.G., Eyre-Walker, A., 2002. Adaptive protein evolution in *Drosophila*. *Nature* 415, 1022–1024.
- Solé-Cava, A.M., Thorpe, J.P., 1991. High levels of genetic variation in natural populations of marine lower invertebrates. *Biol. J. Linn. Soc.* 44, 65–80.
- Stoletzki, N., Eyre-Walker, A., 2007. Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol. Biol. Evol.* 24, 374–381.
- Stoletzki, N., Welch, J., Hermisson, J., Eyre-Walker, A., 2005. A dissection of volatility in yeast. *Mol. Biol. Evol.* 22, 2022–2026.
- Tenaillon, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F., Gaut, B.S., 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc. Natl. Acad. Sci. U. S. A.* 98, 9161–9166.
- Urrutia, A.O., Hurst, L.D., 2001. Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. *Genetics* 159, 1191–1199.
- Vignal, A., Milan, D., SanCristobal, M., Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* 34, 275–305.
- Ward, R.D., Skibinski, D.O.F., Woodward, M., 1992. Protein heterozygosity, protein structure, and taxonomic differentiation. *Evol. Biol.* 26, 73–159.

Chapitre III – Développement des marqueurs moléculaires et génotypage

Tableau 12 : SNP génotypés avec succès (Ø) dans chaque famille F2

Nom du gène	Famille R*O	Famille E*T	Famille Z*AC
Adrenal Gland Protein	Ø	Ø	Ø
AJ557014	Ø		
Amylase Gene A		Ø	Ø
Amylase_GeneB_2		Ø	
Astacin			
ATPase H+	Ø	Ø	
AY321300	Ø	Ø	Ø
Bcl2		Ø	Ø
Bcl2_2		Ø	Ø
BMP	Ø		
BQ426586			
BQ427193			
BQ427367	Ø	Ø	
Calcium Dependant Protein Kinase	Ø		Ø
Cavortine			
Cytochrome P450	Ø	Ø	
Dehydrogenase	Ø		
Drac3		Ø	Ø
Elongation Factor 1 Alpha	Ø	Ø	Ø
Flavin Monooxygenase		Ø	
Flavin Monooxygenase 2		Ø	
Galectine	Ø	Ø	Ø
Glucose 6 Phosphatase		Ø	Ø
Glutaryl coA dehydrogenase	Ø		Ø
Glutathion S transferase	Ø		Ø
Glycogen Phosphorylase			
GlycoProtein Hormone Receptor	Ø	Ø	
Glycogen Synthase	Ø		
HA114		Ø	Ø
HA114_2		Ø	Ø
IK Cytokine	Ø		Ø
IK Cytokine_2		Ø	
Kazal type proteinase inhibitor	Ø		
Laccase		Ø	Ø
Laccase_2	Ø		
Macrophage expressed Protein			
Myosin Light Chain			
Notch homolog 3 variant protein	Ø	Ø	
Paired type Homeodomain Protein			
Plastin	Ø		
Serine	Ø		
Super Oxyde Dismutase		Ø	Ø
sodium/glucose cotransporter	Ø	Ø	Ø
sodium/glucose cotransporter_2	Ø	Ø	
TRAF	Ø	Ø	Ø
Tubulin Alpha	Ø	Ø	Ø
Vasa like			Ø

3.7.4 Génotypage des marqueurs SNP par SNPlex chez les individus de la F2

Le dépouillement des données brutes de génotypage des marqueurs SNP par SNPlex révèle que 77% (47/61) des marqueurs soumis pour génotypage ont pu être génotypés chez les 3 familles F2. Parmi ces 47 EST, 6 ont été génotypés en double (deux sites polymorphes différents) et pour 8 autres EST (Astacin, BQ426586, BQ427193, Cavortine, Glycogen Phosphorylase, Macrophage expressed Protein, Myosin Light Chain et Paired type Homeodomain Protein), aucune donnée n'a été obtenue pour des raisons techniques. Cela ramène le pourcentage de marqueurs exploitables à 54% (33/61-Tableau 12).

3.8 Discussion

3.8.1 Développement de marqueurs SNP

L'identification de SNP, par séquençage partiel d'EST, a été un travail relativement long et fastidieux mais fructueux. Un total de 321 marqueurs a été identifié dans 61 séquences d'EST, soit en moyenne, plus de 5 sites polymorphes par EST. Si le produit de PCR séquencé est celui attendu, la probabilité de révéler des sites polymorphes est donc relativement importante. Seuls, 5% des ESTs ne présentaient aucun polymorphisme. Ce résultat est à mettre en relation avec la taille relativement faible (<200 pb) de ces séquences EST. Plus un amplicon est long, plus la probabilité de trouver du polymorphisme est importante. Le taux de polymorphisme le plus élevé a été estimé à un SNP toutes les 9 paires de bases (partie intronique du gène de l'intégrine). Ce résultat est à relativiser car il aurait été bon de cloner puis de séquencer ce fragment afin de vérifier son exactitude. Cependant, l'ensemble de nos résultats démontre que l'huître creuse est l'un des organismes qui présente un taux de polymorphisme nucléotidique des plus élevés parmi ceux décrits pour le règne animal. Par exemple, chez le tunicier *Ciona savignyi* et le ver rond *Caenorhabditis elegans*, des taux de polymorphismes record ont été mis en évidence avec un SNP toutes les 35 et 20 paires de bases (Small et al., 2007 ; Cutter et al., 2006). Nos résultats confirment de précédents résultats obtenus par Curole et Hedgecock (2005) qui ont estimé le taux de polymorphisme moyen d'un SNP/33 pb chez *C. gigas*.

Par définition, un SNP correspond à un polymorphisme d'une seule base, mais son génotypage dépend principalement des séquences nucléotidiques qui le bordent. Nos résultats révèlent un taux de polymorphisme très élevé : 1 SNP/60pb dans le codant et 1SNP/40pb dans le non codant. Pour la recherche de polymorphisme, cette caractéristique du génome de l'huître est très intéressante mais pose des problèmes pour le génotypage. En effet,

Chapitre III – Développement des marqueurs moléculaires et génotypage

l'abondance d'autres sites polymorphes à proximité du site polymorphe d'intérêt empêche ou restreint le dessin d'amorces spécifiques pour le génotypage. Le site d'intérêt doit être abandonné pour un autre dont le niveau d'information est souvent inférieur. La région promotrice et la région 3' d'un gène sont des régions généralement mieux conservées que la région 5'. C'est pourquoi, il est essentiel de concentrer les efforts de séquençage d'EST et d'identification de SNP dans ces régions.

Le développement de marqueurs SNP est un travail difficilement automatisable à partir du séquençage d'ADN génomique du fait des difficultés de corrections des alignements de séquences et des vérifications des nombreux sites « douteux ». De plus, un grand nombre de ces marqueurs doivent être identifiés pour contrebalancer les « pertes » liées aux difficultés techniques rencontrées pour le génotypage. En effet, parmi les marqueurs initialement identifiés dans les 61 EST, seulement 33 marqueurs (soit 54%) ont pu être intégrés dans la construction de la cartographie génétique (voir chapitre IV). Les SNP sont certes abondants et largement répartis au sein des génomes, mais leur développement en grand nombre reste une tâche qui nécessite beaucoup de temps. L'équilibre entre l'effort de développement et une certaine forme de rentabilité peut être trouvée, en augmentant le nombre d'individus génotypés et réduire ainsi le coût par échantillon de cette étape.

3.8.2 Génotypage des marqueurs microsatellites

Contrairement à l'analyse des SNP, celle des microsatellites est basée sur l'étude de la taille de fragments composés de séquences répétées. Elle est confrontée au phénomène de "bégayement" ou "slippage" de la Taq polymérase lors de l'amplification et de ce fait, souffre des difficultés d'interprétation pouvant être induites par la présence de 'stutter bands'. L'observation des résultats d'électrophorèse lors de l'analyse des marqueurs microsatellites est automatisable, mais la lecture finale du résultat est réalisée et vérifiée manuellement (Coquoz, 2003; Buckleton et al., 2006). De plus, des artéfacts techniques viennent brouiller l'interprétation des résultats. Par exemple, sur un même échantillon d'individus et pour un même marqueur microsatellite, les résultats de deux « runs » de migration peuvent être relativement différents. La fluidité du polymère, support de la migration, et la température ambiante sont des paramètres difficilement contrôlables mais qui ont une nette incidence sur la qualité et la répétabilité des analyses. Une température ambiante élevée tend à fluidifier le polymère, ce qui accélère la vitesse de migration des produits de PCR au cours du « run ». De ce fait, la taille d'un même produit de PCR peut être lue et interprétée différemment (allèles plus ou moins longs). Le génotypage de marqueurs microsatellites a donc été standardisé pour

Chapitre III – Développement des marqueurs moléculaires et génotypage
minimiser les biais techniques et leurs incidences sur la qualité des résultats par le strict respect des conditions de migrations (renouvellement du polymère et température ambiante régulée à 20°C±2°C).

La seconde difficulté du génotypage des marqueurs microsatellites réside dans la correction manuelle des profils de migration. Cette étape importante doit être menée de façon rigoureuse pour différencier les pics qui représentent les deux allèles du marqueur et les pics parasites correspondant à des produits non spécifiques. L'insertion dans le jeu de données de résultats erronés ou manquants a une influence non négligeable sur les étapes suivantes de cartographie génétique (Hackett et Broadfoot, 2003). En effet, au-delà de 4% de données erronées, les distances génétiques, entre marqueurs éloignés d'une distance supérieure à 20 cM, tend à être sous évaluée.

3.8.3 Biais d'usage du code génétique

Le biais d'usage du code génétique dans un génome est la résultante des effets de la sélection, des mutations et de la dérive génétique (Bulmer, 1991). Nos résultats montrent qu'il existe un biais mutationnel chez l'huître creuse qui tend à enrichir son génome en bases A et T. De plus, l'analyse du biais d'usage du code génétique chez l'huître creuse a mis en évidence la corrélation entre la fréquences des codons optimaux et le niveau d'expression des gènes. En conséquence, le biais d'usage du code génétique chez *C. gigas* est paradoxalement moins équilibré. Dans les gènes qui sont peu exprimés, ceux-ci sont pauvres en G et C alors que dans les gènes fortement exprimés le taux de G et C apparaît supérieur. Ce résultat souligne que l'analyse du biais d'usage du code est difficile à établir sans une évaluation des niveaux d'expression (Duret et Mouchiroud, 1999). Dans le cas du génome de *C. gigas*, un effet de sélection agit vraisemblablement au niveau du mécanisme de traduction. Cette sélection favoriserait le maintien des codons qui se terminent par un G ou un C, en dépit de la pauvreté du génome de l'espèce en bases G et C. Cependant, l'efficacité de la sélection n'est pas suffisante pour donner une forte impulsion à l'enrichissement au contenu en G et C, car le taux de GC dans certains gènes reste relativement faible. Nos résultats sont similaires à ceux récemment mis en évidence chez le nématode *C. elegans* (Marais et al., 2001) dont le génome est pauvre en GC où la sélection traductionnelle tend à favoriser codons se terminant par G ou C (Cutter et al., 2006).

3.9 Perspectives concernant le développement de marqueurs moléculaires chez *C. gigas*.

De récents travaux d'analyse d'EST chez *Crassostrea virginica* (Quilang et al., 2007) ont révélé un grand nombre de SNP (>4500). De plus, la publication d'un grand nombre 13000 séquences EST chez quatre espèces de bivalves marins, dont l'huître creuse (Tanguy et al., 2008), constitue une source supplémentaire où de nouveaux marqueurs SNP pourront être identifiés et intégrés à notre carte consensus. Dans la perspective de la recherche de QTL, des travaux sont en cours pour identifier du polymorphisme nucléotidique parmi des gènes candidats (Huvet et al., 2004). De plus, l'identification de gènes candidats, issus de l'analyse de micro-arrays et PCR en temps réel, est actuellement en cours, suite aux travaux menés chez *Crassostrea gigas* (Fleury et al., 2008). Ces travaux ont plus particulièrement identifiés 104 EST candidats impliqués dans la réponse au phénomène de mortalité estivale.

Suite aux très récents travaux menés en collaboration avec le Génoscope, plus de 50000 EST sont aujourd'hui disponibles chez l'huître creuse. A partir de ces données, une détection de marqueurs *in silico* a été menée. Par l'utilisation de moteurs de recherche spécialisés : SSR Finder (www.maizemap.org/bioinformatics/SSRFINDER/) et PupaSuite, (Conde et al., 2006), 232 marqueurs microsatellites et 7530 SNP ont pu être identifiés. Ces résultats soulignent l'intérêt des EST comme source de développement de marqueurs génétiques. Néanmoins, ces marqueurs, obtenus par une approche *in silico*, doivent encore être validés par le génotypage d'individus de familles de référence ou issus de populations naturelles afin d'évaluer leur intérêt. Les prochaines cartes génétiques vont permettre de développer l'utilisation de ces ressources conjointement à l'évolution des techniques de génotypage à très haut débit.

A moyen terme, l'élaboration d'un jeu de marqueurs SNP, utilisables pour différentes applications telles que la génétique des populations ou la cartographie de liaison, peut être envisagée. En effet, les techniques récentes de génotypage à haut et très haut débit telles que les technologies Illumina®, Affimetrix®, Sequenom® ou 454 de Roche deviennent accessibles car leur coût diminue et un nombre croissant de plateformes de génotypage s'équipent.

CHAPITRE IV

CONSTRUCTION D'UNE
CARTOGRAPHIE GENETIQUE
CHEZ *C. GIGAS*

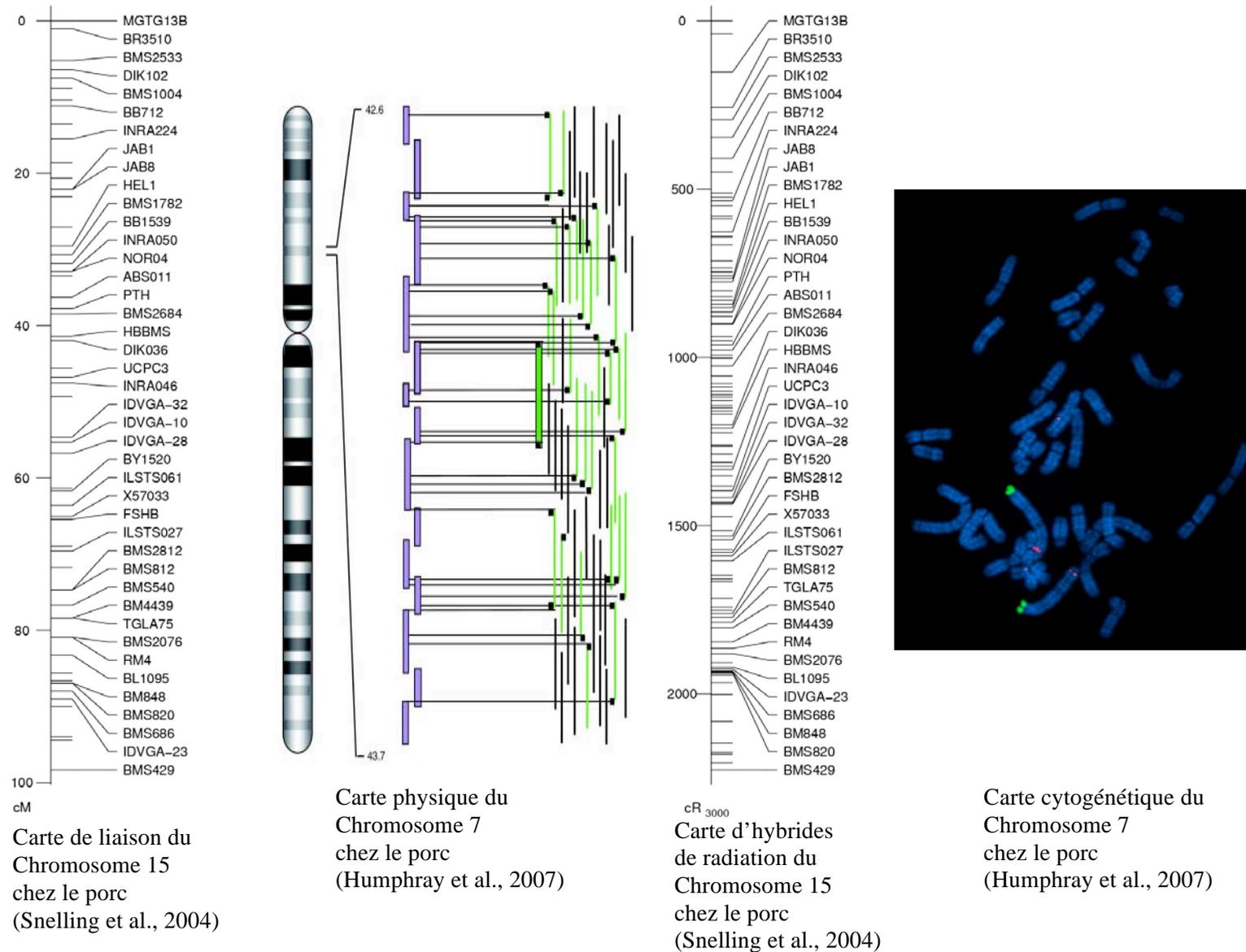


Figure 27 : Différents types de cartes (liaison, physique, d'hybride de radiation et cytogénétique)

Il existe deux types différents de cartes génétiques (Figure 27) : les cartes de liaison et les cartes physiques. La carte de liaison est une représentation statistique du génome en fonction du taux de recombinaison entre marqueurs moléculaires constituant des groupes de liaisons. Une carte physique est l'ordonnement de marqueurs le long de chromosomes. Il existe différents types de cartes physiques. Ainsi, la carte cytogénétique obtenue par hybridation moléculaire *in situ* sur métaphases colorées (FISH) permet de localiser et d'ordonner directement des sondes correspondant à des gènes, des marqueurs génétiques, etc. La carte d'hybrides de radiation (RH) repose, quant à elle, sur la fréquence de cassures chromosomiques induites artificiellement, observées chez des hybrides d'irradiation (ex : cellules humaines lourdement irradiées et fusionnées à des cellules de hamster). En ce qui concerne les cartes de liaison, que nous appellerons cartes génétiques dans ce chapitre par simplification, des marqueurs moléculaires de différentes natures (allozymes, AFLP, microsatellites, SNP, etc...) peuvent être combinés au sein d'une même cartographie génétique. L'intégration de marqueurs microsatellites reste très intéressante car ce sont des marqueurs très polymorphes qui jouent le rôle de marqueurs d'ancrage (Schlötterer et Goldstein, 1999).

4.1. Définition et Principe

La carte génétique est l'assemblage en groupes de liaison et l'ordonnement de marqueurs moléculaires grâce à l'analyse statistique de leur ségrégation au cours des générations. Les distances génétiques se mesurent en centimorgan (cM). Cette unité est définie comme la distance entre deux loci chromosomiques pour lesquels un produit méiotique sur cent est recombinant (voir encadré 3). A priori, les recombinaisons se produisent aléatoirement le long des chromosomes au cours de la méiose. Plus la probabilité de recombinaison entre deux marqueurs est élevée, plus la distance entre ceux-ci est grande. Par conséquent, la cartographie génétique consiste à suivre la ségrégation de loci au sein de la descendance. Le niveau de résolution de la cartographie dépend directement de la densité de marqueurs et de la taille et de la structure de la population analysée. Ainsi, trois niveaux d'analyse peuvent être effectués : (1) l'analyse de ségrégation, (2) l'analyse des individus dits recombinants et (3) l'analyse du déséquilibre de liaison (de Vienne, 1998). L'analyse de la ségrégation permet la construction d'une carte globale du génome en étudiant la transmission allélique au sein de familles informatives. Les analyses d'individus recombinants et du déséquilibre de liaison permettent de travailler sur des distances génétiques plus courtes et ainsi affiner la détection de QTL. La construction d'une carte génétique s'effectue classiquement en plusieurs étapes :

4.1.1 Les tests de liaison entre les loci

Le plus souvent, l'existence d'une liaison est vérifiée par l'analyse du rapport de vraisemblance connu sous le nom de **LOD Score** pour **Logarithm of the Odd ratio** (Morton, 1955). Celui-ci se décompose comme suit :

Soit H0 : il y a une liaison génétique entre 2 locus ($r < 0,5$) ; la vraisemblance de cette hypothèse est notée $e^L(r = \theta)$

Soit H1 : il n'y a pas de liaison génétique entre ces 2 locus ($r = 0,5$) ; la vraisemblance de cette hypothèse est notée $e^L(r = 0,5)$

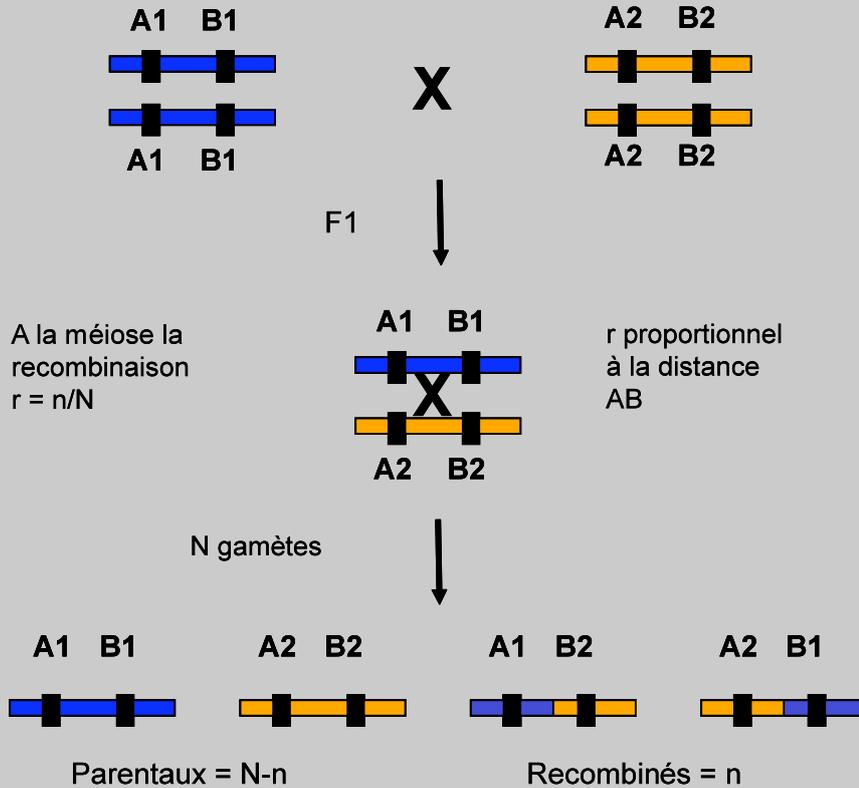
$$\text{d'où LOD score} = \text{Log}_{10} \frac{e^L(r = \theta)}{e^L(r = 0,5)}$$

Lorsque le LOD score est supérieur à une valeur seuil, il existe une liaison génétique. Par exemple, lorsque le LOD = 3, cela signifie que l'hypothèse de liaison est 1000 fois plus vraisemblable que l'hypothèse d'indépendance des 2 loci.

4.1.2 Le calcul des distances génétiques

Il existe des régions du génome qui recombinent beaucoup plus qu'attendu en fonction de leur distance physique (points chauds de recombinaison Brown et al., 1991). Inversement, il existe de très longues portions d'ADN dans des régions centromériques et télomériques où le phénomène de recombinaison est réprimé par la présence d'hétérochromatine (Tanksley et al., 1992) ou d'ADN hautement répété. Ces zones montrent par conséquent des distances génétiques plus faibles entre les marqueurs qui tendent à être regroupés en "cluster". Si le rapport entre distance physique et distance génétique varie au niveau intra-chromosomique, il varie plus généralement lorsque l'on compare des espèces à contenu d'ADN différents. A titre d'exemple, 1 cM équivaut en moyenne à 13 ; 3,5 ; 2,5 et 0,23 millions de paires de bases

Encadré 3: principe de la cartographie génétique et calcul du taux de recombinaison entre locus



Au cours de la méiose, qui donne lieu à la formation des gamètes, des échanges entre les chromosomes homologues se produisent : il s'agit du « crossing-over », représentés schématiquement entre deux locus A et B. Le calcul de la fréquence de cette recombinaison (r) permet d'estimer la distance entre loci. Cette fréquence sera d'autant plus élevée que les deux locus seront éloignés l'un de l'autre sur le chromosome. r est estimé comme le rapport entre les gamètes de type recombiné et ceux de type parental. Si $r = 0$, les locus sont totalement liés et aucun gamète de type recombiné se produit lors de la méiose. Si $0 < r < 0,5$, les deux locus sont liés. La valeur de r est une mesure de cette liaison et est estimée par la proportion de gamètes recombinés. Enfin, si $r = 0,5$, les deux locus ségrègent de façon indépendante, soit ils se trouvent sur deux chromosomes différents, soit sur le même chromosome mais à une distance suffisante pour qu'il y ait autant de gamètes recombinés que de gamètes parentaux.

respectivement chez le pin maritime, le blé, le maïs et l'arabette des dames. Ainsi, des distances qui peuvent sembler faibles en terme de distance génétique s'avèrent parfois importantes en terme de distance physique. Le taux de recombinaison entre marqueurs identiques est également variable selon l'environnement (Allard, 1956; Towe et al., 1964; Lu, 1974; Ihrke et al., 1975; Rose et al., 1979; Zhuchenko et al., 1983; Tulsieram et al., 1992a), le fond génétique (Beavis et al., 1991) et le sexe (Moran et al., 1983; Robertson, 1984; Vizir et al., 1990; de Vicente et al., 1993; Paterson et al., 1991; Graner et al., 1991; Eshed et al., 1992; Groover et al., 1995). Des croisements inter-spécifiques donnent également des longueurs de carte plus faibles que des croisements intra-spécifiques (Rick, 1969; Tadmor et al., 1987; Douches et al., 1988; Gebhardt et al., 1991; Causse et al., 1995). Un mauvais appariement des chromosomes homologues à la méiose plus ou moins important en fonction du degré de divergence des espèces parentales entraînerait une baisse de la fréquence des chiasma et se traduirait par un taux de recombinaison plus faible entre les marqueurs (Borts et al., 1987).

La première fonction de cartographie a été formulée par Sturtevant en 1913 qui établit que le taux de recombinaison R (voir encadré 3) est égale à une unité de distance génétique M . Cette hypothèse est partiellement vraie car elle ne s'applique qu'aux petites distances génétiques. Au delà, les taux de recombinaison ne sont pas additifs pour des fréquences de recombinaison importantes ($>0,2$). La conversion du taux de recombinaison en distances génétiques nécessite une fonction de conversion mathématique complexe destinée à modéliser l'interférence entre la recombinaison. Il existe quatre fonctions cartographies développées respectivement par Haldane (1919), Kosambi (1944), Carter et Falconer (1951) et Morgan (1914). Celles-ci se distinguent entre elles principalement par la prise en compte de l'interférence entre crossing-overs. L'interférence se définit comme la probabilité que deux crossing-over se produisent proches l'un de l'autre. Par exemple, la fonction de Morgan est basée sur une interférence totale.

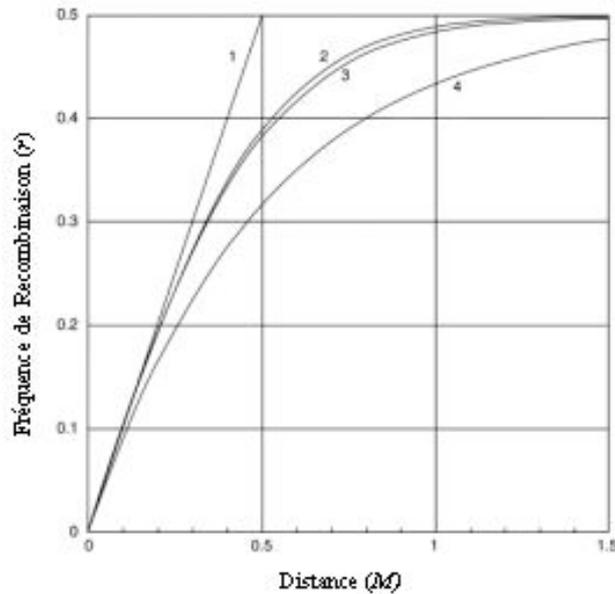


Figure 28 : Fonctions de cartographies linéaires (1) Fonction de Sturtevant's ($R = M$) ; (2) Fonction de Haldane avec interférence (3) Fonction de Kosambi ; (4) Fonction de Haldane sans interférence

Cependant, la fonction mathématique la plus couramment utilisée pour le calcul des distances est la fonction de Kosambi (1944; Figure 28). Ce calcul de distance prend donc en compte les crossing-over multiples ainsi que l'interférence. Si les crossing-over sont indépendants, il n'y a pas d'interférence. Celle-ci s'écrit ainsi (où r est la fréquence de recombinaison) :

$$d (M) = 1/4 \ln [(1 + 2r)/(1 - 2r)]$$

Les fonctions de calcul des distances génétiques convertissent des fréquences de recombinaison en distances génétiques mais elles n'estiment pas les distances physiques entre les marqueurs. Les régions où la recombinaison est relativement forte (Hot-spot), le ratio entre distance génétique (cM) et distance physique (kilobases) sera important. Cependant, lorsque la recombinaison est uniforme le long du chromosome, une fonction de conversion recombinaison-distance génétique peut aider à déterminer la localisation d'un gène sur une carte physique. L'intervalle de confiance de la position du marqueur moléculaire cartographié peut être estimé par de nombreux algorithmes. Cependant, une formule relativement simple est appliquée. Cette formule, établie chez la souris, calcule l'erreur standard selon la formule suivante : $[r(1-r)/n]^{1/2}$ où r est la taux de recombinaison et n le nombre de population.

Ensuite, les cartographies génétiques peuvent être comparées entre elles : il s'agit de la cartographie comparée qui est généralement réalisée entre deux cartes établies chez deux espèces différents (ex : homme/souris) ou chez différentes espèces appartenant au même genre (ex : Salmonidés, Gharbi et al., 2006) pour évaluer le niveau d'homologie entre ces cartes. Cela peut orienter le choix d'un gène candidat qui pourrait être intéressant de cartographier chez une espèce proche. Enfin, lorsque des cartographies génétiques sont disponibles chez des espèces relativement proches (du même genre, par exemple), il est alors possible de mesurer le degré de synténie qui sépare ces deux espèces. Ce degré de synténie reflète la conservation de l'ordre des gènes des chromosomes entre ces deux mêmes espèces (Lyons et al., 2008).

4.1.3 Test de conformité aux lois de Mendel

La ségrégation est un concept fondamental en génétique. Il correspond à la séparation des deux allèles d'un même locus au cours de la formation des gamètes chez un individu diploïde. Cela assure en théorie la représentation de ces deux allèles dans la descendance avec un ratio 1:1. Lorsque la transmission des allèles ne respecte plus ce ratio, il y a une distorsion de la ségrégation. C'est pourquoi, en cartographie génétique, il est essentiel de tenir compte de ce biais en testant l'hypothèse nulle H_0 : la ségrégation des marqueurs est mendélienne (ou non distordue). La plupart des tests de conformité repose sur l'hypothèse d'absence de distorsion de ségrégation (Whitkus, 1998). Cette distorsion peut avoir plusieurs origines et engendrer d'importantes conséquences notamment en cartographie génétique.

- **Le biais méiotique** : au cours de la méiose, la moitié des gamètes (spermatozoïde ou ovocyte) reçoit un exemplaire du gène paternel et l'autre moitié des gamètes reçoit un exemplaire du gène maternel. Cette répartition équitable des allèles n'est cependant pas toujours de règle, la méiose constituant un moment favorable à l'apparition de conflits génétiques. En effet, les deux formes alléliques au niveau d'un locus sont en compétition pour leur représentation dans les gamètes. Dans une cellule hétérozygote, si l'un des allèles a le pouvoir d'entraîner un biais dans la ségrégation méiotique, cela lui permet d'être surreprésenté dans la population de gamètes. La distorsion de ségrégation méiotique est universellement répandue et a été mise en évidence dès 1957 par Sandler et Novitski chez de nombreuses espèces d'animaux et de plantes (Sapienza et al., 1982). Alors que les divisions méiotiques sont

symétriques chez le mâle (autant de chromatides que de produits de méiose), chacune de ces deux divisions sont asymétriques chez la femelle (ovocyte et globules polaires ; figure 29). En effet, la séparation non-aléatoire des chromatides entre l'ovocyte et les globules polaires résulte en une représentation inégale des allèles (Pardo-Manuel de Villena et al., 2001). Ce phénomène rejoint la notion de taux de transmission de la distorsion ou TRD (Transmission Ratio Distorsion) qui peut se définir comme un « écart statistique significatif aux règles de ségrégation Mendélienne attendues quelle que soit la cause » (Pardo-Manuel de Villena et al., 2000).

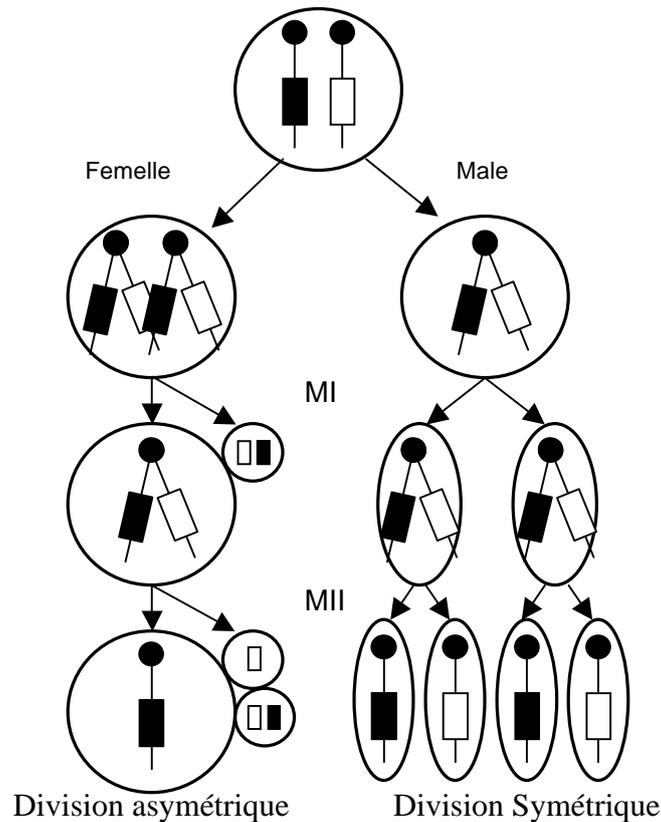


Figure 29: Schématisation du phénomène de biais méiotique (Pardo-Manuel de Villena et al., 2001).

- **Les réarrangements structuraux :** des mécanismes en amont de la fécondation, tels que les translocations (ex : translocation robertsonienne chez l'homme) et la perte de chromosomes (aneuploïdie) sont à l'origine de réarrangement structuraux qui affectent la viabilité des gamètes, notamment chez les végétaux (Faure et al., 1993).
- **La sélection en phase zygotique :** suite à la fécondation, les premières divisions cellulaires forment un zygote ou œuf. Chez les espèces à très forte fécondité, telle que l'huître creuse, une partie, parfois importante, des zygotes ne survivent pas au delà de ce stade de développement (prédation, virus, interactions génotype-environnement défavorable). La distorsion résulte alors du différentiel de taux de survie chez ces zygotes qui appartiennent à différentes classes gamétiques (Falconer et al., 1996).

Diverses études démontrent l'effet de cette distorsion sur les tests de liaison entre loci et les estimations des distances génétiques. En effet, une importante distorsion augmente le nombre de faux-positifs dans les tests de liaison et tend à surévaluer les distances génétiques (Garcia-Dorado et al., 1992; Lorieux et al., 1995a; Lorieux et al., 1995b). Cependant, la plupart des méthodes statistiques utilisées pour construire des cartes génétiques ne tiennent pas compte de ce phénomène (Lander et al., 1987; Jiang et al., 1997). Zhu et collaborateurs (2007) proposent une méthode statistique relativement complexe de reconstruction des cartes de liaison établies à partir de marqueurs dits « distordus ». Les résultats obtenus montrent que la carte corrigée permet une détection de QTL(s) plus juste. De manière plus pratique, il est également possible de minimiser l'effet de la distorsion de ségrégation en travaillant sur des individus échantillonnés aux stades les plus précoces de leur cycle de vie. Cette stratégie a été appliquée avec succès par Hubert et Hedgecock (2004) chez *Crassostrea gigas* dans le cadre de l'établissement de la première carte génétique basée sur des marqueurs co-dominants. En effet, les auteurs ont échantillonné des larves d'huître âgées de 11 jours car, à ce stade, la sélection en phase zygotique est encore faible.

4.2 Cartographies génétiques établies chez les principales espèces d'intérêt aquacole

De part les importants intérêts commerciaux, de nombreuses cartographies ont été établies chez les espèces aquacoles et agricoles. La littérature relative à ce sujet est très importante, c'est pourquoi, nous ne passeront en revue ici que les études de cartographies publiées récemment chez les principales espèces aquacoles : le zebrafish (Kelly et al., 2000),

tilapia (Lee et al., 2005), le poisson chat ou « channel catfish » (Waldbieser et al., 2001; Liu et al., 2003), « Japanese flounder » (Castano-Sanche et al., 2007), le bar (Chistiakov et al., 2005 ; Volckaert et al., 2007), la carpe (Liao et al., 2007), la daurade (Franch et al., 2006) et les salmonidés (Young et al., 1998; Sakamoto et al., 2000; Nichols et al., 2003; Guyomard et al., 2006). Des cartographies génétiques ont également été établies chez les ostréidés (Yu et al., 2003; Li et Guo, 2004; Hubert et Hedgecock., 2004; Lallias et al., 2007), les pectinidés (Li et al., 2005; Qin et al., 2007) et autres espèces telles que la moule bleue (Lallias et al., 2007), l'ormeau (Sekino et al., 2007) ou encore chez les différentes espèces de crevettes (Wilson et al., 2002; Li et al., 2003; Zhang et al., 2007). Le développement de cartographies génétiques est relativement récent chez les espèces aquacoles et connaît un intérêt croissant (figure 30). L'ensemble des caractéristiques de ces cartographies est résumé dans le tableau 13, pages 144 à 148.

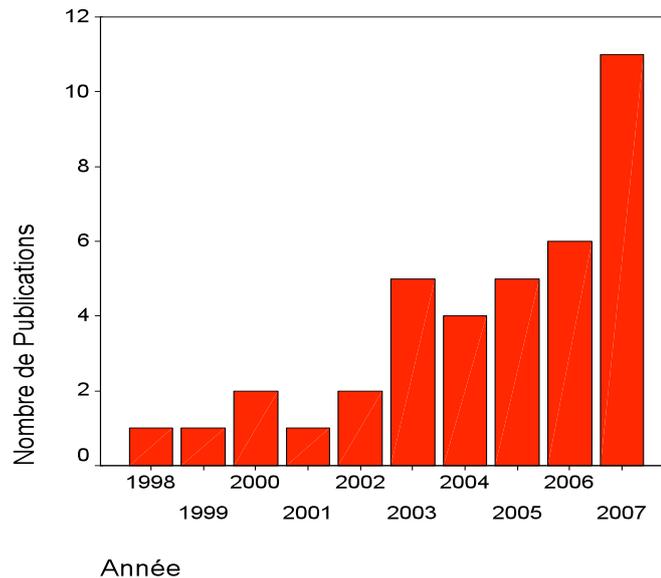


Figure 30 : Evolution du nombre de publications de cartes génétiques chez les espèces d'intérêt aquacole au cours de la dernière décennie (nombre de publications identifiées par comptage sur la base de publication Pubmed en mars 2008).

Chez l'huître creuse, deux cartes de liaison ont été publiées. La première, développée en 2004 par Li et collaborateurs, se compose de 11 groupes et comporte 119 marqueurs AFLP espacés en moyenne de 9,5 cM. La seconde, développée en 2004 par Hubert et Hedgecock, repose sur 102 marqueurs microsatellite, et présente onze groupes de liaison chez le mâle et 12 chez la femelle couvrant une distance respective de 616 et 770 cM. Pour cette seconde carte, la distance moyenne séparant deux marqueurs est de 10,5 cM chez le mâle et 13,6 cM chez la femelle.

Chapitre IV – Cartographie Génétique

Tableau 13 : Cartographies publiées chez les espèces d'intérêt aquacole

Espèces	Caractéristiques de la carte	Nombre et type de marqueurs	Design expérimental	Logiciel utilisé	Références bibliographique associées
Zebrafish - <i>Dario rerio</i>	Carte consensus: 25 LG 1364 Marqueurs Longueur 3011cM	642 SSCP, 593 Microsatellites et 129 gènes	2 familles F1 hybrides et 42 individus double hybrides	MapManager/MapMaker	(Kelly et al., 2000)
Tilapia - <i>Oreochromis niloticus</i> (n=22)	Carte consensus: 24 LG Longueur 1311cM Intervalle 2,4cM	525 microsatellites et 21 gènes	1 famille F2 obtenue par le croisement interspécifique de <i>O. niloticus</i> et <i>O. aureus</i>	Crimap	(Lee et al., 2005)
Barbue de rivière - <i>Ictalurus punctatus</i> (n=29)	Carte consensus: 32 LG 262 Marqueurs Longueur 1958cM Intervalle 8,7cM	293 Microsatellites	2 familles de référence	Crimap	(Waldbieser et al., 2001)
Japanese Flounder - <i>Paralichthys olivaceus</i> (n=23)	Carte mâle: 25 LG 223 Marqueurs Longueur 741,1cM Intervalle 8cM Carte femelle: 27 LG 294 Marqueurs Longueur 670,4cM Intervalle 6,6cM	82 Microsatellites 149 AFLP 101 Microsatellites 203 AFLP	Croisement inter-lignée (mâle produit par gynogenèse)	MapManager	(Coimbra et al., 2003)
Bar - <i>Dicentrarchus labrax</i> (n=25)	Carte mâle: 25 LG Longueur 567,4cM Intervalle 3,5cM Carte femelle: 25 LG Longueur 905,9cM Intervalle 5,59cM	162 Microsatellites 11 SNP	1 famille F1	Crimap	(Chistiakov et al., 2005)

Chapitre IV – Cartographie Génétique

Espèces	Caractéristiques de la carte	Nombre et type de marqueurs	Design expérimental	Logiciel utilisé	Références bibliographique associées
Carpe - <i>Cyprinus carpio</i> (n=50)	Carte consensus: 50LG 268marqueurs Longueur 4111cM	110 microsattellites 105 gènes et 57 RAPD	1 famille haploïde obtenue par gynogenèse	MapManager	(Sun et al., 2004)
Daurade - <i>Sparus aurata</i>	Carte consensus: 26 LG 204Marqueurs Longueur 1241,6cM	204 microsattellites	1 famille F1	Crimap	(Franch et al., 2006)
<i>Salmo trutta</i> (2n=80)	Carte mâle: 20 LG 301 marqueurs Longueur 346,4cM Carte femelle: 20LG 301 marqueurs Longueur 912,5cM	288 microsattellites et 13 allozymes	2 mâles de population méditerranéenne croisés en backcross par des femelles atlantiques	Carthagene	(Gharbi et al., 2006)
Saumon Atlantique - <i>Salmo salar</i> (n=40)	Carte mâle: 31 LG 251 Marqueurs Longueur 103cM Carte femelle: 33LG 230 Marqueurs Longueur 901cM	473 AFLP et 54 microsattellites	2 familles de pleins frères (full-sib)	Calcul des « lod scores » pour chaque paires de marqueurs informatifs puis JoinMap	(Moen et al., 2004)
Truite Arc en Ciel - <i>Oncorhynchus mykiss</i> (n=29-32)	Carte consensus: 31 LG Longueur 2750cM Intervalle 4cM	796 Microsattellites et 2 SNP	2 familles F2	Carthagene	(Guyomard et al., 2006)
Huître creuse du Pacifique - <i>Crassostrea gigas</i> (n=10)	Carte mâle:10LG 96 marqueurs Longueur 758,4cM Intervalle 8,8cM Carte femelle: 11 LG 119 marqueurs Longueur 1030,7cM Intervalle 9,5cM	349 AFLP	2 familles F1 (dont une backcross)	Mapmaker	(Li et Guo, 2004)

Chapitre IV – Cartographie Génétique

Espèces	Caractéristiques de la carte	Nombre et type de marqueurs	Design expérimental	Logiciel utilisé	Références bibliographique associées
Huître creuse du Pacifique - <i>Crassostrea gigas</i> (n=10)	Carte mâle: 10 LG 88 marqueurs Longueur 616,1cM Intervalle 8cM Carte femelle: 10 LG 86 marqueurs Longueur 770,5cM Intervalle 10,4cM	115 microsatellites	3 familles F2	Mapmaker	(Hubert et Hedgecock, 2004)
Huître creuse américaine - <i>Crassostrea virginica</i> (n=10)	Carte mâle: 12 LG 114 marqueurs Longueur 647cM Intervalle 6,3cM Carte femelle: 12 LG 84 marqueurs Longueur 904,3cM Intervalle 12,6cM	153 AFLP 3 microsatellites et 2 EST 129 AFLP 3 microsatellites et 1 EST	1 famille F1	Mapmaker	(Yu et al., 2003)
Huître plate européenne - <i>Ostrea edulis</i>	Carte parent 1: 9LG 104 marqueurs Longueur 471,2cM Intervalle 4,86cM Carte parent 2: 10LG 117 Marqueurs Longueur 450cM Intervalle 4,21cM	235 AFLP et 16 microsatellites	1 famille F2	Crimap/Mapmaker	(Lallias et al., 2007)
Moule Bleue - <i>Mytilus edulis</i>	Carte mâle: 14LG 116 marqueurs Longueur 825cM intervalle 8cM Carte femelle: 14LG 121 marqueurs Longueur 862,8cM Intervalle 8cM	791 AFLP	1 famille F1	MapMaker	(Lallias et al., 2007)

Chapitre IV – Cartographie Génétique

Espèces	Caractéristiques de la carte	Nombre et type de marqueurs	Design expérimental	Logiciel utilisé	Références bibliographique associées
Zhikong scallop - <i>Chlamys farreri</i> (n=19)	Carte mâle: 23 LG 166 Marqueurs Longueur 2468cM Intervalle 14,9cM	603 AFLP	1 famille F1 issue du croisement de deux individus provenant d'une population Japonaise et Chinoise	MapMaker	(Wang et al., 2005)
	Carte femelle: 25LG 198 marqueurs Longueur 3130cM Intervalle 14,6cM				
	Carte Mâle: 20 Lg 197 marqueurs Longueur 1630,7cM Intervalle 9,2cM	667 AFLP	1 Famille F1 issue du croisement de deux individus provenant d'une population Japonaise et Chinoise	MapMaker	(Li et al., 2005)
	Carte Femelle: 19LG 166 marqueurs Longueur 1503,9cM Intervalle 10,2cM				
Ormeau - <i>Haliotis discus hannai</i> (2n=36)	Carte mâle: 18LG 167 marqueurs Longueur 702,4cM Intervalle 4,7cM	180 Microsatellites	3 familles F1	JoinMap	(Sekino et al., 2007)
	Carte femelle: 19LG 160 marqueurs Longueur 888.1cM Intervalle 6,3cM				
Pacific White Shrimp - <i>Litopenaeus vannamei</i>	Carte mâle: 45LG 267Marqueurs Longueur 3220cM	118 AFLP et 30 Microsatellites	200 familles de plein-frères (full-sib)	MapMaker	(Zhang et al., 2007)
	Carte femelle: 45LG 319 Marqueurs Longuer 4134cM				

Chapitre IV – Cartographie Génétique

Espèces	Caractéristiques de la carte	Nombre et type de marqueurs	Design expérimental	Logiciel utilisé	Références bibliographique associées
Kuruma prawn - <i>Penaeus japonicus</i> (n=43)	Carte mâle: 43LG 227 marqueurs Longueur 1781cM Intervalle 9,7cM <hr style="width: 50%; margin-left: 0;"/> Carte femelle: 31 LG 125 marqueurs Longueur 1026cM Intervalle 10,9cM	401 AFLP	1 famille F2	MapMaker	(Li et al., 2003)
Black Tiger Shrimp - <i>Penaeus monodon</i> (n=44)	Carte consensus: 19LG 63 marqueurs Longueur 1412cM Intervalle 22cM	673 AFLP	3 familles F2	MapMaker	(Wilson et al., 2002)
White shrimp - <i>Penaeus vannamei</i> (n=44)	Carte mâle: 47LG 182 marqueurs Longueur 2116cM Intervalle 15,6cM <hr style="width: 50%; margin-left: 0;"/> Carte femelle: 51Lg 212 marqueurs Longueur 2771cM Intervalle 17,1cM	741 AFLP	1 famille F2 issue d'un stock domestiqué après 5 générations de sélection	MapManager QTX	(Perez et al., 2004)

4.3 Matériel et Méthodes

4.3.1 Matériel biologique et marqueurs moléculaires

Notre schéma expérimental est basé sur l'analyse de la ségrégation chez trois familles F2 d'individus au stade naissain (voir chapitre II) et les données de génotypage de marqueurs SNP et microsatellites (voir chapitre III) chez deux groupes d'individus de phénotypes extrêmes (individus morts au cours des premiers jours du suivi et individus survivants). Les microsatellites issus de la publication de Hubert et Hedgecock (2004) ont été sélectionnés selon leur répartition, afin d'assurer une couverture du génome la plus large possible tout en réduisant le nombre de marqueurs à géotyper et la masse de travail en biologie moléculaire.

4.3.2 Choix du logiciel de cartographie

Les analyses de liaison ont suscité le développement de nombreux logiciels de cartographie génétique. Une revue non exhaustive est disponible sur le site internet suivant : <http://linkage.rockefeller.edu/soft/>. La plupart de ces logiciels utilisent les mêmes statistiques de base pour le test de liaison entre marqueurs. Leurs principales différences résident dans le type de populations et de croisements (Backcross, F2, etc...), leur plateforme de fonctionnement et interface (Windows, Unix), le codage des données et les options qu'ils proposent (analyse multipoint, recherche d'erreurs dans le jeu de données, modules de représentation graphique des cartes, etc...). Quatre logiciels sont plus couramment utilisés : Crimap (Lander et al., 1987), Mapmaker (Lander et al., 1987), JoinMap (Stam, 1993) et Carthagene (Schiex et al., 1997). Ces quatre logiciels sont couramment utilisés chez les espèces d'intérêt aquacoles comme l'huître plate européenne *Ostrea edulis* (Lallias et al., 2007) ; la moule bleue *Mytilus edulis* (Lallias et al., 2007) ; l'huître creuse *Crassostrea gigas* (Hubert et al., 2004) ou la truite arc en ciel *Oncorhynchus mykiss* (Guyomard et al., 2006).

Pour établir une carte consensus à partir du jeu de données obtenus sur les familles F2, nous avons utilisé le logiciel Crimap. Ce dernier est un logiciel relativement robuste où les différentes options proposées permettent notamment de tester l'ordre des marqueurs au sein des groupes de liaison (Lander et al., 1987).

4.3.3 Principe d'utilisation du logiciel Crimap

Le logiciel de cartographie Crimap (Lander et al., 1987) a été développé pour réaliser la construction de groupes de liaison à partir de données de génotypes. Son principe de fonctionnement est relativement simple. Les données brutes sont regroupées dans un fichier d'extension « .gen ». Lorsque ce fichier est complet, celui-ci contient les informations sur les 3 générations d'individus (F0, F1 et F2), leurs ascendants directs, leur sexe, et leur génotype aux marqueurs moléculaires associés. Chaque allèle d'un marqueur microsatellite a été codé selon la taille du fragment. Pour les marqueurs de type SNP, les deux allèles ont été codés en données binaires (1/2). Enfin, les données manquantes ont été codées par un zéro.

Dés lors que le fichier de données est en forme, la fonction « prepare » du logiciel est utilisée afin de créer quatre fichiers d'extensions différentes. Premièrement, le fichier « .par » contient l'ensemble des paramètres de l'analyse (ordre de marqueurs, analyse sexe spécifique, etc...). Deuxièmement, le fichier « .loc » contient les informations relatives à chacun des marqueurs (nombre de méioses informatives, etc...). Troisièmement, le fichier « .dat » contient les données réorganisées selon son propre format alors que le quatrième fichier « .ord » contient l'ordre des marqueurs.

Ensuite, la première étape vers la construction de la carte génétique consiste en l'exécution de la commande 'Twopoint'. Cette commande effectue les tests de liaison pour chaque paire de marqueurs et leur attribue un LOD score supérieur ou égal à 3 en fonction de l'importance ou de la force de la liaison. Un Lod score égal à 3 signifie que l'hypothèse de liaison est 1000 fois supérieure à l'hypothèse inverse car $\text{Log } 3=1000$. Ce Lod score minimum peut être modifié afin d'éliminer les scores qui reflètent de faux positifs, c'est à dire qui mettent en évidence une liaison génétique entre deux loci alors que celle ci n'existe pas. Généralement, un faux positif se repère par un score proche de trois. Prenons un exemple : soit cinq loci A, B, C, D et E. Les couples de loci A/B et C/D sont liés entre eux par un Lod score très élevé (ex : supérieur à 20). E est lié à A et B par un Lod score élevé (env. 10) et à D par un Lod score proche du seuil de signification (ex: 3,02). Si E ne montre aucune liaison avec C ou un autre marqueur du groupe auquel appartiennent C et D, alors la liaison entre E et D, sera considérée comme un faux positif. Ainsi, le marqueur E sera placé dans le groupe de liaison des loci A et B. Le paramètre de seuil minimal de Lod score peut donc être utilisé dans le but d'éliminer les faux positifs et de lever les ambiguïtés d'assignation de loci au sein des

groupes de liaison. Par association, il est alors possible de construire l'ensemble des groupes de liaison.

Lorsque tous les groupes de liaison sont formés, les analyses suivantes sont menées séparément et de manière identique sur chacun des groupes de liaison. Il convient alors de déterminer la distance qui sépare les marqueurs et leur ordre le plus probable dans le groupe de liaison. Pour cela, la fonctionne BUILD est exécutée. Néanmoins, pour éviter les erreurs dans cette étape de construction et éliminer les distances génétiques aberrantes, la commande FLIP doit être également utilisée. Cette commande recalcule les distances génétiques et l'ordre des marqueurs au sein du groupe en intervertissant de deux à (n-1) marqueurs et attribue un score à chacun des alignements alternatifs. Si le score est positif, cela signifie que l'alignement alternatif de deux ou (n-1) marqueurs est plus probable que l'ordre actuel. Il convient donc de conserver l'ordre le plus probable des marqueurs du groupe de liaison. Généralement l'ordre le plus probable correspond au groupe de liaison le plus court. Dès lors, tout marqueur dont la position reste ambiguë a été écarté des analyses suivantes. De plus, par l'exécution de la commande CHROMPIC, qui permet de mettre en évidence sur chacun des chromosomes, les éventuelles erreurs dans l'estimation de la fréquence de recombinaison ont été mises en évidence.

4.3.4 Application à nos jeux de données

Le fichier initial d'extension « .gen » soumis pour analyse sous Crimap contenait les informations relatives aux individus des trois familles F2, soit :

- 10 individus F0 (Y1, K, H1, AC, Q, R, R1, Z, Q1, et V),
- 6 Individus F1 (R4, O4, E2, T2, Z3 et AC3),
- 300 individus par famille F2 x3 = 900 (96 individus morts au cours de l'événement de mortalité estivale en 2006 et 96 ayant survécu à ce même épisode),
- La liste complète des noms des marqueurs moléculaires tous types confondus,
- Les données de génotypes pour les deux types de marqueurs moléculaires :
 - 47 marqueurs SNP ont été génotypés chez les 300 individus F2 de chacune des familles F2 ainsi que *de novo* chez les individus F0 et F1
 - 67 marqueurs microsatellites dont 49 marqueurs précédemment publiés (nommés UcdCg117 par exemple) et 18 marqueurs développés *in silico* (CgSili 3 à 57) chez des individus issus de population naturelle.

Le jeu de données final regroupe environ 110.000 génotypes, obtenu chez 1000 individus environ, et a été déposé dans son intégralité sur la base de données « Respecies » (<http://www.respecies.org/>) abritée par le Roslin Institute. La proportion de données manquantes est estimée à dix sept pourcents. En effet, pour l'ensemble des marqueurs microsatellites et quelle que soit la famille F2, seulement 2 groupes de 96 individus par condition (individus morts/vivants) ont été génotypés afin de réduire la masse de travail de biologie moléculaire. Cette réduction du nombre d'individus génotypés se justifie par le fait que les microsatellites sont des marqueurs moléculaires relativement informatifs (cf. chapitre III). De plus, les individus échantillonnés correspondent aux extrémités de la courbe de distribution car notre schéma expérimental est basé sur l'analyse des phénotypes les plus divergents (voir chapitre V – génotypage sélectif).

Les paramètres d'analyse de liaison et de construction des groupes de liaison sont les suivants : seuls les couples de marqueurs dont la liaison est supérieure ou égale à trois ont été retenus pour l'analyse dite en « Twopoint ». Dès lors, par association, les marqueurs ont été regroupés puis les distances génétiques entre marqueurs adjacents ont été estimées selon la formule de Kosambi (1944). Enfin, l'ordre des marqueurs de chacun des groupes a été testé par la commande 'flipsn'. L'ordre le plus probable a été conservé et les distances génétiques recalculées si nécessaire.

4.3.5 Représentation graphiques des cartes

Le logiciel de cartographie Crimap ne permet pas de représenter graphiquement les cartes. L'utilisation d'un logiciel tiers est nécessaire. A partir de l'ordre des marqueurs et des distances génétiques qui les séparent, le logiciel Mapchart (Voorrips, 2002), met en forme les cartes de liaison. Son utilisation est relativement facile. Le fichier de construction de la carte se compose d'une séquence de groupes de liaison. Chaque groupe est caractérisé par une première ligne de texte spécifiant le nom de ce groupe et d'une succession de lignes portant le nom des locus et leur distance sur la carte.

4.3.6 Comparaison avec la carte génétique précédente établie par Hubert et Hedgecock chez *Crassostrea gigas*

Une comparaison directe de l'organisation de la carte consensus et de la carte génétique publiée par Hubert et Hedgecock (2004) a été effectuée. Le test du rhô de Spearman a été utilisé pour étudier les corrélations entre la longueur de chacun de ces groupes, le nombre de marqueurs par groupe de liaison et les distances moyennes inter-marqueurs. De plus, ce même test statistique a été utilisé pour évaluer la corrélation entre la distance qui sépare chaque paire de marqueurs communs entre les deux cartes.

4.4 Résultats

4.4.1 Distorsion de ségrégation

L'analyse par χ^2 des fréquences alléliques a révélé d'importantes distorsions de ségrégation quel que soit le type de marqueur. En effet, 51, 48 et 21% des marqueurs montrent une distorsion significative au seuil 5% respectivement parmi les familles, E*T, Z*AC et R*O. Cette distorsion est répartie de manière aléatoire entre les groupes de liaisons et la direction de cette distorsion est variable. Cependant, celle-ci se caractérise par un déficit en hétérozygote ($P = 12,41$; $ddf = 1$; $P = 0,05$). L'analyse de la distorsion de ségrégation montre que la direction de cette distorsion varie le long des groupes de liaisons. Dans ces régions, les fréquences d'hétérozygotes sont significativement plus basses que la fréquence théorique ($P = 12,41$; $ddf = 1$; $P = 0,05$).

4.4.2 Carte génétique consensus

La carte consensus, qui regroupe les données obtenues chez les 3 familles F2, repose sur quatre vingts marqueurs dont 29 SNP, 13 microsatellites développés *in silico* et 38 microsatellites précédemment publiés (Li et al., 2003; Yamtich et al., 2005; Yu et al., 2007). Le nombre de groupes de liaison identifiés est en accord avec le nombre de chromosomes observés en phase haploïde chez l'huître creuse, soit 10 (tableau 14, figure 31). Les groupes de liaison ont été nommés de façon identiques à ceux publiés par Hubert et Hedgecock (2004) pour faciliter leur comparaison (voir paragraphe 4.4.3). Ceci a été possible car 35 marqueurs microsatellites sont communs aux deux études.

Chapitre IV – Cartographie Génétique

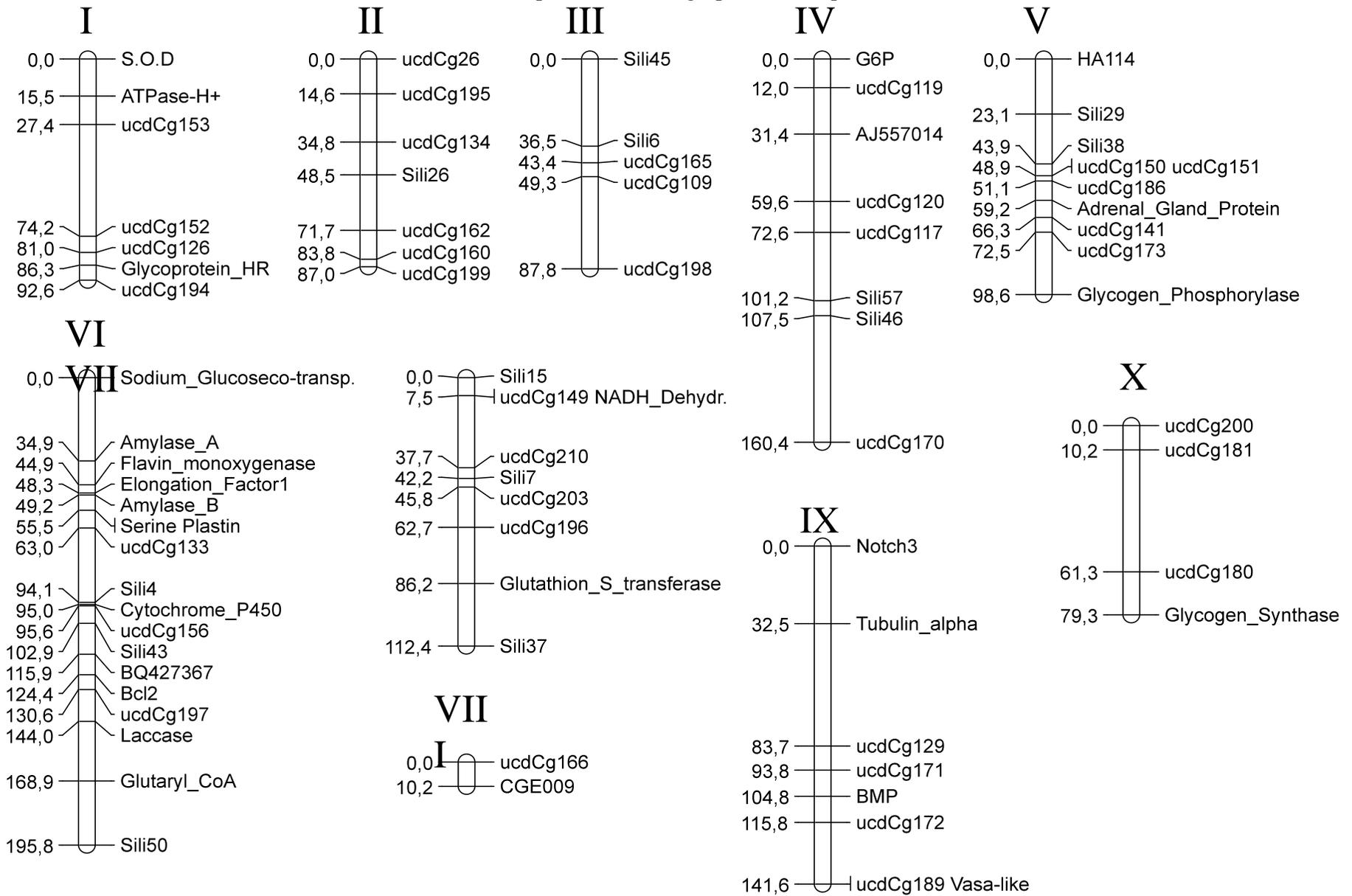


Figure 31 : cartographie génétique consensus composée de 10 groupes de liaison, comportant 80 marqueurs et d'une longueur totale de 1062cM

Tableau 14 : Longueur, intervalle minimum, moyen et maximum et constitution des groupes de liaison.

Groupe	Nbr de marqueurs (SNP/Msat)	Longueur Totale	Intervalle Moyen (cM)	Intervalle Maximum (cM)	Intervalle Minimum (cM)
LG1	7 (3/4)	92,6	13,2	46,8	5,3
LG2	7 (0/7)	87	14,0	23,2	12,1
LG3	5 (0/5)	87,8	17,6	38,5	5,9
LG4	9 (3/6)	160,4	20,1	52,7	6,3
LG5	10 (3/7)	98,6	9,9	26,1	0
LG6	19 (13/6)	195,8	11,0	34,9	0
LG7	9 (2/7)	112,4	12,5	30,2	0
LG8	2 (0/2)	10,2	10,2	10,2	0
LG9	8 (4/4)	141,6	17,7	51,2	0
LG10	4 (1/3)	79,2	19,8	51,2	10,2
Total	80 (29/51)	1065,6	13,3	52,7	0

Cette carte consensus s'étend sur une distance génétique de 1062,4 cM avec le plus grand groupe de liaison qui mesure 195,8 cM alors que le plus petit d'entre eux, ne mesure que 10,2 cM. L'intervalle moyen entre deux marqueurs adjacents est compris entre 9,9 et 20,1 cM et les intervalles minimum et maximum sont respectivement compris entre 0 et 12,1 cM et 10,2 et 52,7 cM. La taille théorique du génome de l'huître creuse a été estimée selon la formule de Chakravarti (1991) où la longueur observée de chacun des groupes de liaison a été multipliée suivant la formule : $L = (m+1)/(m-1)$ où m est le nombre de marqueurs composant le groupe de liaison. Ainsi, la taille théorique a été estimée à 1390 cM ce qui signifie que la carte consensus couvre environ 76,4% du génome.

Le plus petit groupe de liaison (10,2 cM) est composé de 2 marqueurs alors que le plus grand des groupes de liaison (195,8 cM), regroupe 19 marqueurs dont 13 SNP. Une analyse de la corrélation par le calcul du Rho de Spearman confirme la corrélation entre le nombre de marqueurs de chacun des groupes de liaison et sa longueur ($\rho=0,884$; $P>0,001$ – figure 32) ce qui est généralement constaté dans la construction des cartes de liaison.

Les 80 marqueurs cartographiés représentent 70,2% des 114 marqueurs génotypés et déposés dans la base de données « Resspecies ». Cependant, parmi ces 114 derniers, 8 marqueurs SNP et 6 microsatellites étaient monomorphes chez les trois familles F2. De plus, deux sites polymorphes par EST ont été génotypés pour 6 ESTs. Cela ramène le nombre de marqueurs informatifs à 94, ce qui signifie que 85% (80/94) des marqueurs ont été cartographiés. Les 14 marqueurs (15%) manquants correspondent soit à des marqueurs pour lesquels aucune liaison pour un Lod score significativement supérieur ou égal à deux n'a été

mise en évidence (Msat : *Ucdcg193* et SNP : *Drac3*) ; soit à des marqueurs (n=12) introduisant des distances aberrantes (généralement de 100cM) lors de la construction des groupes de liaison consensus (*Ucdcg 45, 145, 157 et 191, Microsatellites in silico 3, 32, 39, 44 et 56* et les SNP Calcium dépendant Kinase, IK-Cytokine et TRAF).

Cependant, pour limiter l'effet de la distorsion de ségrégation mise en évidence précédemment, les analyses ont été effectuées sur un jeu de données modifié. En effet, les données de génotypes présentant le plus de distorsion pour chacun des marqueurs n'ont été utilisées. Par exemple, si les génotypes du marqueur *UcdCg117* présentaient des niveaux de distorsion différent chez les 3 familles F2, les données de génotypes présentant une distorsion élevée ($\text{Chi}^2 > P$) n'ont pas été intégrées pour le calcul des distances génétiques puis l'établissement des cartes consensus et sexe spécifique.

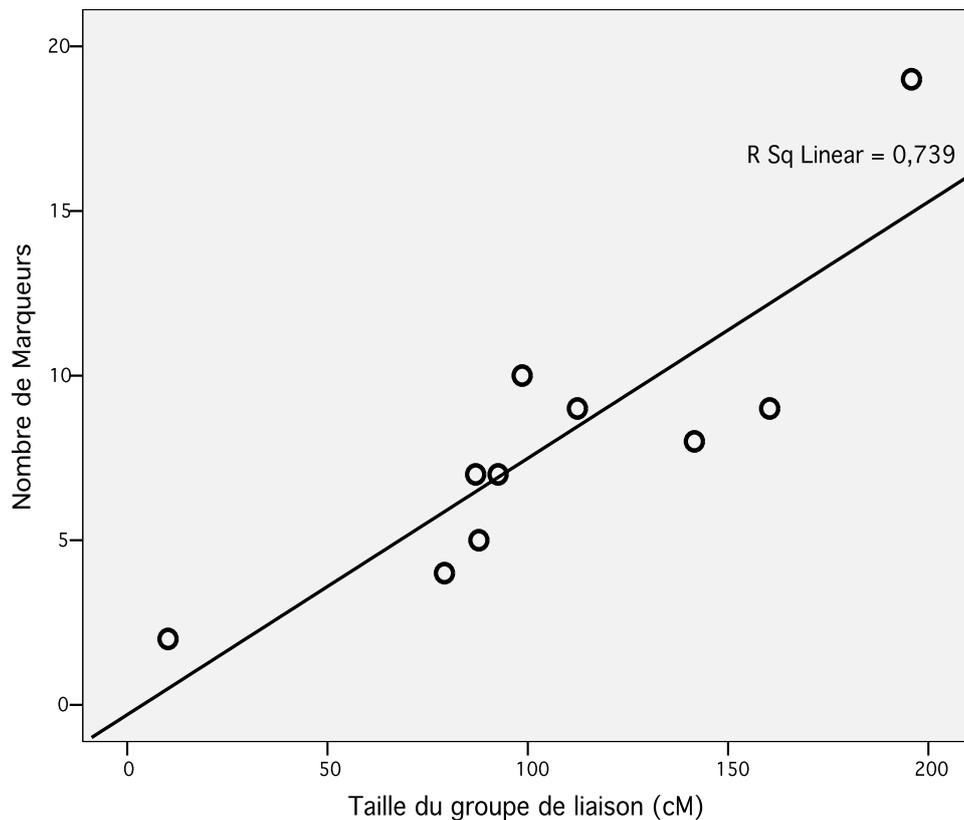


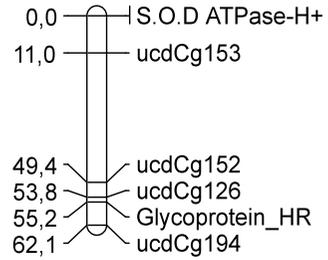
Figure 32 : Corrélation entre le Nombre de marqueurs et la Taille du groupe de liaison (cM)

4.4.3 Cartes consensus sexe spécifique

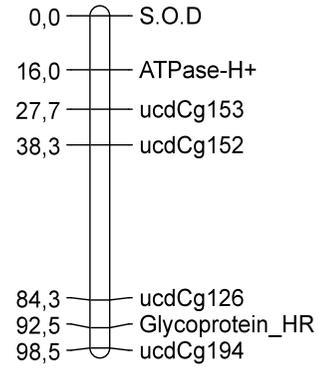
Les cartographies génétiques consensus mâle et femelle ont été établies selon le même protocole que celui utilisé pour la construction de la cartographie génétique consensus et en utilisant le même jeu de données tronqué. Les cartes femelle et mâle regroupent chacune 79 marqueurs (Figure 33). Celles-ci diffèrent uniquement par deux marqueurs situés dans les groupes de liaison IV et VI (microsatellite *ucdCg170* absent chez la femelle et le gène de la Sodium Glucose co-transporter absent chez le mâle). Ces deux cartes s'étendent sur des distances totales respectives de 1075 et 782 cM (femelle/mâle) et l'intervalle moyen entre deux marqueurs est estimé à 13,8 et 10,5 cM (femelle/mâle). La corrélation entre le nombre de marqueurs et la longueur respective des groupes de liaison est significative pour les deux cartes ($\rho=0,876$; $P>0,01$). Le ratio entre la distance génétique totale de la carte femelle et la distance génétique de la carte chez le mâle a été estimée à 1,37 car le taux de recombinaison était supérieur chez la femelle que chez le mâle. Les groupes de liaison I à III et VI à IX sont plus grands chez la femelle que chez le mâle. Cependant, la différence de taille entre groupes de liaison homologues chez la femelle et le mâle est relativement variable avec un ratio minimum de 1,1 (LG II et VI) et un ratio maximum de 4,9 (LG IX). Enfin, la couverture du génome est relativement similaire quel que soit le sexe (76,2% chez la femelle contre 75,9% chez le mâle).

Chapitre IV – Cartographie Génétique

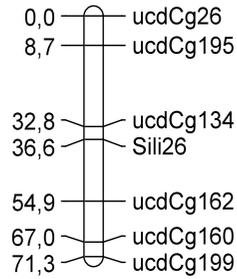
I_Male



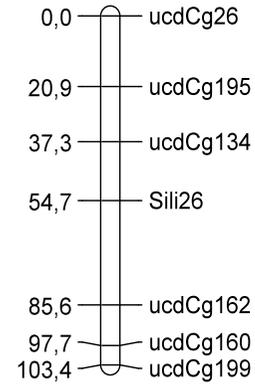
I_Femelle



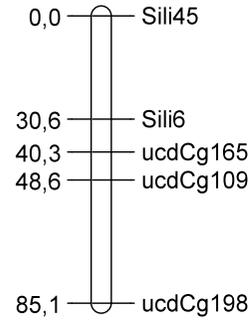
II_Male



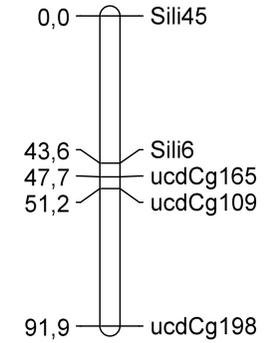
II_Femelle



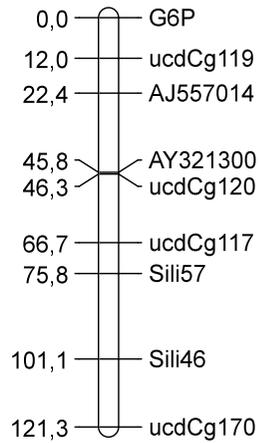
III_Male



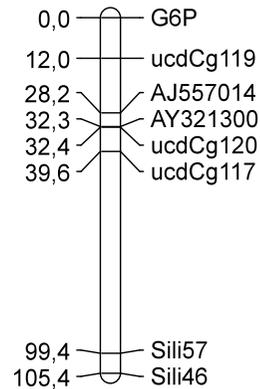
III_Femelle



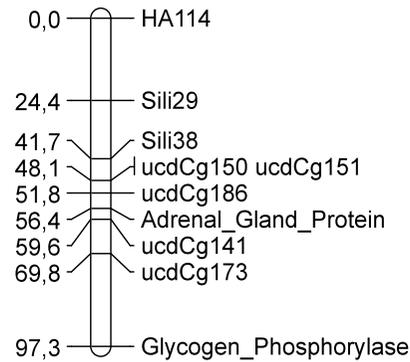
IV_Male



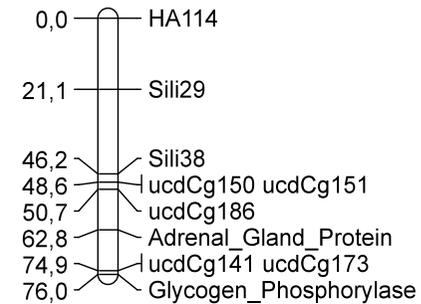
IV_Femelle



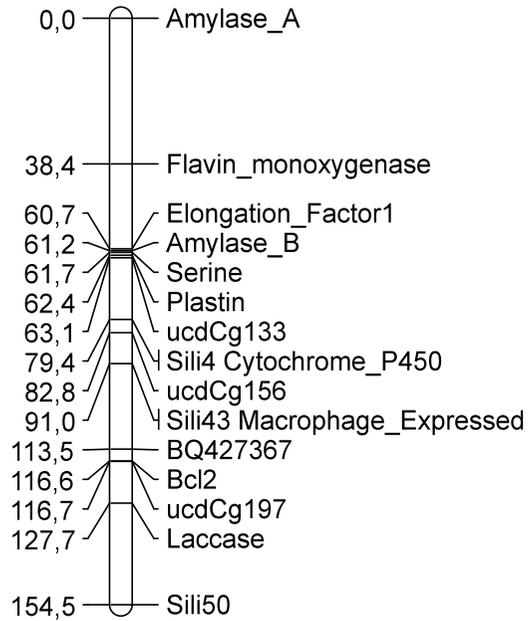
V_Male



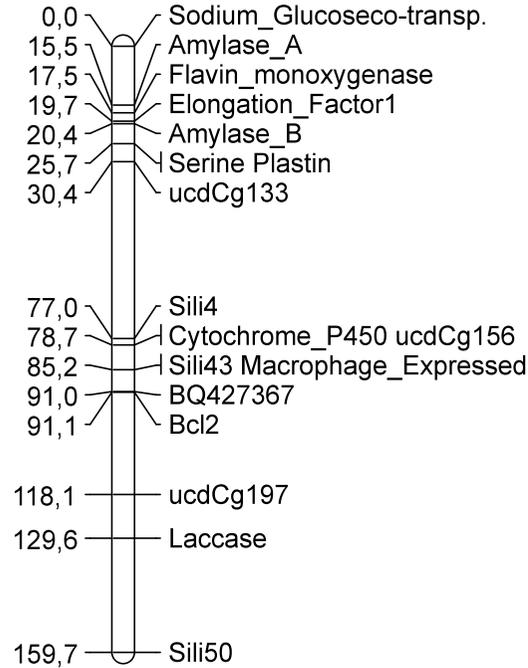
V_Femelle



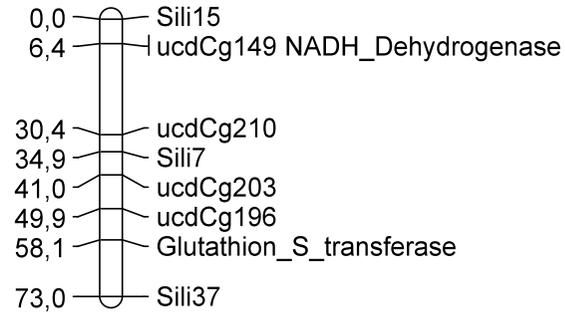
VI_Male



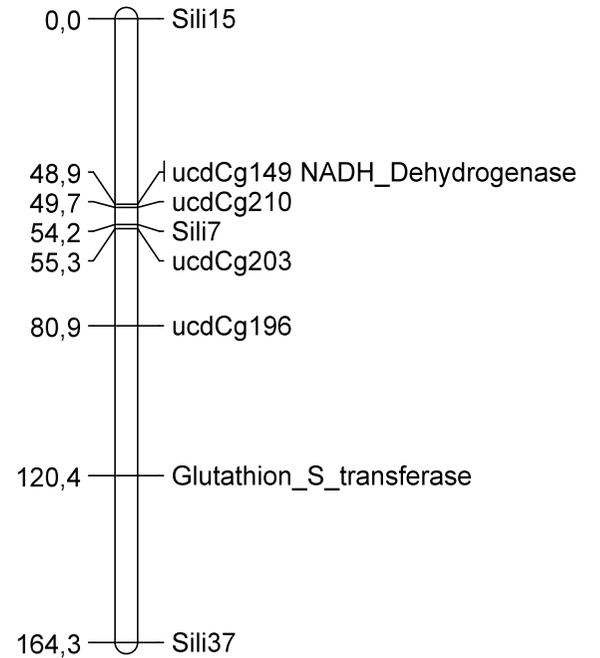
VI_Femelle



VII_Male



VII_Femelle

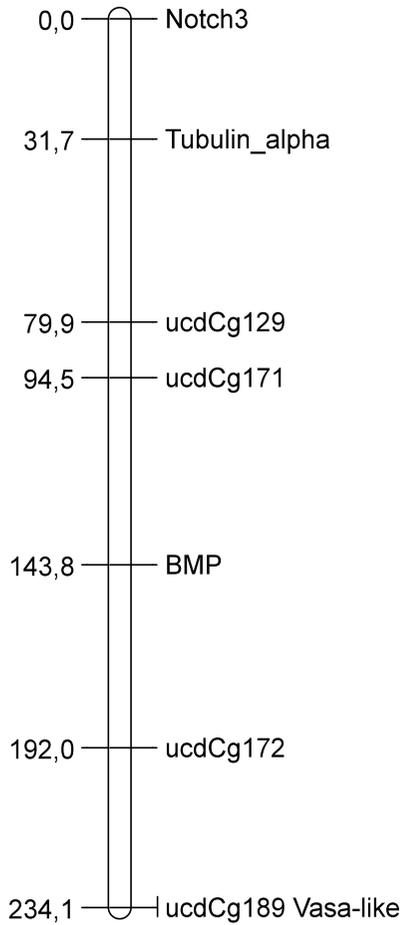
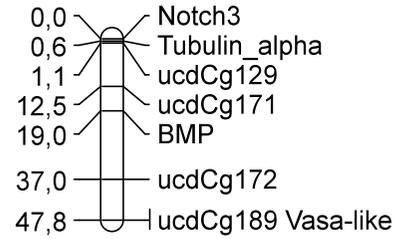
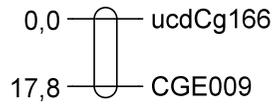


VIII_Male

VIII_Femelle

IX_Male

IX_Femelle



X_Male

X_Femelle

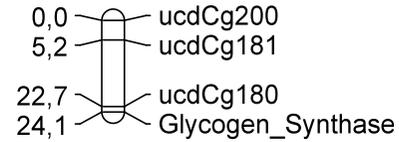
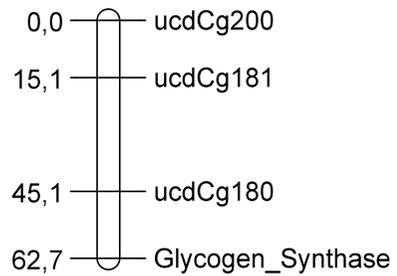


Figure 33 : Cartes génétiques consensus mâle et femelle

4.4.4 Comparaison avec la carte génétique développée par Hubert et Hedgecock

La carte génétique consensus, publiée par Hubert et Hedgecock (2004), est construite à partir de marqueurs microsatellites. Il apparaissait relativement intéressant de dresser un comparatif en terme de longueur de groupes de liaison, d'ordre et de distance entre marqueurs, entre la carte consensus que nous avons établie et cette précédente carte. Le tableau 15 réunit les données respectives pour ces deux cartes.

Tableau 15 : comparatif du nombre de marqueur, de la longueur et de l'intervalle pour chaque groupe de liaison de *Crassostrea gigas*

Groupe	Nbr de marqueurs/LG		Longueur des LG (cM)		Intervalle moyen (cM)	
	Présents travaux	Hubert et Hedgecock	Présents travaux	Hubert et Hedgecock	Présents travaux	Hubert et Hedgecock
LG1	7	20	92,6	< 147,9	13,2	> 7,4
LG2	7	11	87	< 108,3	14	> 9,8
LG3	5	8	87,8	< 131,1	17,6	> 16,4
LG4	9	12	160,4	> 96	20,1	> 8
LG5	10	11	98,6	> 84,6	9,9	> 7,7
LG6	19	7	195,8	> 87,9	11	< 12,6
LG7	9	10	112,4	> 60,4	12,5	> 6
LG8	2	5	10,2	< 50,4	10,2	= 10,1
LG9	8	6	141,6	> 31,6	17,7	> 5,3
LG10	4	2	79,2	> 19,5	19,8	> 9,8

Entre les deux cartes génétiques, la longueur des groupes de liaison est le seul paramètre corrélé ($\rho = 0,6$; $P > 0,05$).

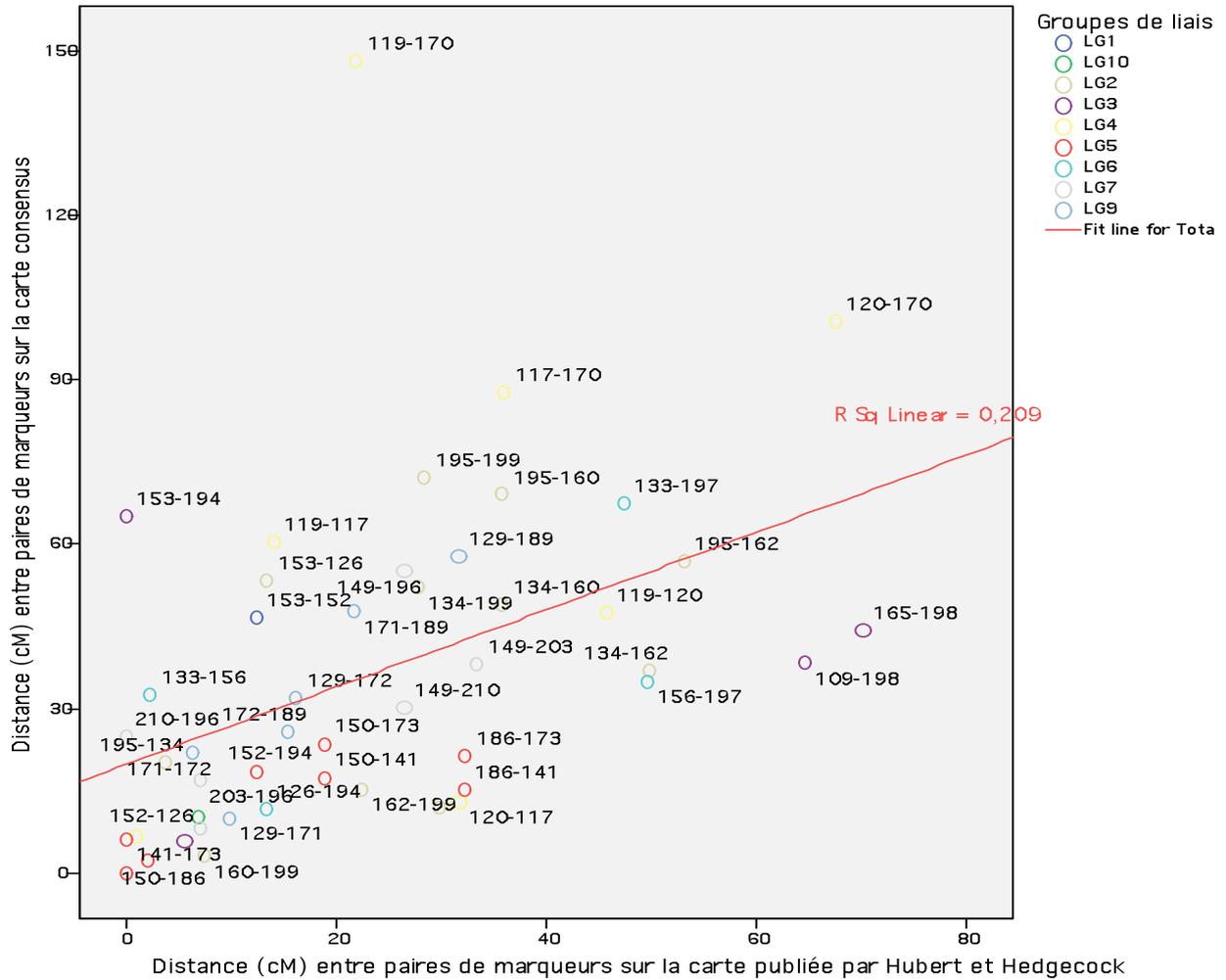


Figure 34 : Corrélation entre les distances inter-marqueur entre paires de marqueurs communes aux carte consensus et carte publiée par Hubert et Hedgecock (2004). Chaque couple de marqueurs est annoté selon le numéro des microsatellites (ex : couple *UcdCg119-UcdCg170* annoté 119-170).

De plus, la corrélation de la distance inter-marqueur de chaque paires de marqueurs communes entre la carte consensus et la carte publiée par Hubert et Hedgecock est significative ($\rho = 0,564$; $P > 0,01$ -Figure 34). Un seul point se situe clairement en dehors du nuage de points. Ce point, qui appartient au LG4, correspond au couple commun de marqueurs *UcdCg170* et *UcdCg119* qui sont séparés de 148,4 cM sur la carte consensus et séparés de 22 cM sur la carte précédemment publiée par Hubert et Hedgecock.

4.5 Discussion

4.5.1 Facteurs influençant la construction des cartes génétiques

Plusieurs paramètres peuvent affecter sur la construction des cartes génétiques : la ségrégation de distorsion, la taille de la (ou des) population(s) génotypée(s), le pourcentage de données manquantes, les erreurs de génotypage et enfin la qualité de l'ADN. L'effet de chacun de ces paramètres a été minimisé autant que possible afin d'établir une carte génétique consensus la plus probable.

4.5.1.1 Distorsion de ségrégation

La distorsion de ségrégation est très souvent rencontrée dans la construction de cartographies génétiques et les analyses QTL (Lyttle, 1991). Il est alors souvent difficile de traiter des jeux de données affectées par ce phénomène (Luo et al., 2005). C'est pourquoi, nous avons essayé de minimiser son impact en réduisant le jeu de données par l'élimination, pour chaque marqueur, des données de génotypes montrant la plus grande ségrégation de distorsion dans une ou deux familles F2. L'analyse systématique, par un test du Chi², des proportions alléliques de chacun des marqueurs moléculaires utilisés dans la construction de la carte consensus démontre le biais induit par la distorsion de ségrégation. Ce phénomène est décrit très largement dans la littérature notamment chez les espèces présentant un fardeau génétique important telle que l'huître creuse (Launey et Hedgecock, 2001; Hubert et Hedgecock, 2004). C'est pourquoi, la création du fichier de données regroupant l'ensemble des données les moins distordues de chacun des marqueurs, chez chacune des trois familles ségrégeantes F2, a notamment permis un gain de précision dans le calcul des distances génétiques et l'ordonnement des marqueurs au sein des groupes de liaison. En effet, la construction de la carte de liaison consensus avec le jeu de données au complet n'a pas aboutit à une carte regroupant un maximum de marqueurs, respectant le nombre de chromosomes et surtout présentant des distances inter-marqueurs cohérentes. La carte consensus, construite à partir du jeu de données tronqué, présente une taille similaire à celle précédemment publiée par Hubert et Hedgecock (2004). De plus, un nombre supérieur de marqueurs, notamment SNP (12 vs 29), ont été incorporés dans cette nouvelle carte consensus, ce qui porte le nombre total de marqueurs à 80.

L'analyse de la distorsion de ségrégation a montré que la direction de cette distorsion varie le long des groupes de liaisons. Dans ces régions, les fréquences d'hétérozygotes sont

significativement plus basses que la fréquence théorique. Deux hypothèses, relativement liées, permettent d'expliquer ce phénomène. Tout d'abord, la première hypothèse repose sur le taux de consanguinité. Plus il est important chez une population, plus la fréquence d'une classe génotypique (généralement hétérozygote) est sous représentée. Cela aboutit à la production d'un stock de gamètes dont le niveau d'homozygotie est important. Cela est notamment vrai chez des individus ou populations impliquées dans des programmes de sélection (Schwarz-Sommer et al., 2003). La seconde hypothèse repose sur les taux de survie en phase larvaire chez l'huître creuse qui sont généralement de l'ordre de quelques pourcents. De ce fait, ce phénomène tend à augmenter le taux d'homozygotie chez la descendance et peut modifier le contrôle de la fitness des individus (Zhu et al., 2007; Luo et al., 2005) réduisant par voie de conséquence la viabilité des descendants.

Par comparaison avec la carte génétique publiée par Hubert et Hedgecock (2004), la distorsion de ségrégation observée dans notre jeu de données, même tronqué, est supérieure avec respectivement 21, 48 et 51% de marqueurs distordus dans chacune des familles F2 contre 2, 2 et 4% de marqueurs distordus chez les trois familles F2 génotypées dans leur expérience. Les pourcentages de marqueurs distordus sont très bas chez les familles F2 génotypées par Hubert et Hedgecock car les auteurs ont choisi d'échantillonner des larves onze jours après la fécondation, diminuant ainsi l'impact de la sélection. Dans notre expérience, la construction de la carte de liaison aurait du être construite également à partir d'individus échantillonnés au stade larvaire. Cela présente l'avantage de minimiser le biais de ségrégation de distorsion, mais dans notre cas, nous obligeait à génotyper des individus au stade larvaire pour construire une carte de liaison et répéter cette étape de génotypage, pour les mêmes marqueurs, chez les individus échantillonnés au cours de l'expérience de phénotypage. Notre priorité étant la recherche de QTL liés aux caractères de survie et de résistance à l'infection par le virus OsHV1, notre choix s'est porté sur la construction de la carte génétique à partir de jeunes individus (naissain).

4.5.1.2 Qualité de l'ADN extrait

Lors du suivi de mortalité en raceway, les individus morts ont été prélevés chaque jour. Néanmoins, la dégradation tissulaire chez des animaux, de taille comprise entre 2 et 7 cm, est relativement rapide. En effet, il n'était pas rare de retrouver uniquement les deux valves d'un animal dépourvues de tissus. La dégradation des tissus est une des causes probables de la distorsion de ségrégation. En effet, si l'ADN extrait sur des individus morts est de moindre qualité, cela aurait pu affecter de manière significative la détection d'allèles lors du génotypage. Ainsi, un individu hétérozygote pour un locus peut être lu comme homozygote et un individu homozygote peut être lu comme un individu n'apportant aucune donnée. Ce type d'erreur est généralement qualifié sous le nom d'allèle 'Drop-out'. C'est pourquoi, l'ADN de chacun des individus morts a migré sur gel d'agarose (1,5%) et a été dosé individuellement par spectrophotométrie pour qu'à l'issue de l'extraction, la quantité d'ADN extraite soit supérieure ou égale à $100 \text{ ng} \cdot \mu\text{L}^{-1}$, ce qui permet, en théorie, la réalisation de nombreuses réactions d'amplification et réduit le risque d'allèle 'Drop out'.

4.5.1.3 Taille de la population, pourcentage de données manquantes et erreurs de génotypage.

La création du fichier de données tronqué n'a pas réduit le nombre d'individus par famille mais cela a réduit la proportion de données manquantes et/ou fausses à 4% et la proportion de données présentant de la distorsion. Selon Hackett et Broadfoot (2003), ce pourcentage est tout à fait acceptable. En effet, ces auteurs ont mené différentes simulations à partir d'un même jeu de données en y introduisant, tour à tour, une distorsion de ségrégation, un taux d'erreur et de données manquantes plus importants. Les variations de ces paramètres ont été mesurées par comparaison avec une carte génétique de référence. Leurs résultats révèlent qu'en deçà de 4% de génotypes manquants et/ou d'erreurs de génotypage les distances génétiques et l'ordonnement des marqueurs n'est pas affecté de manière significative. Les simulations montrent que plus la taille des populations étudiées est petite, plus les effets sur les cartes génétiques sont importants. Cependant, si l'espace moyen entre deux marqueurs est de 15 cM et que la taille de la population est supérieure à 150 individus, de faibles fréquences d'erreurs n'affectent pas la longueur de la carte (Shields et al., 1991). Chacune des trois familles F2 se compose de 300 individus, ce qui, au regard des simulations menées par (Hackett et Broadfoot, 2003), renforce la robustesse de la cartographie consensus.

Au cours du processus de construction des groupes de liaison, douze marqueurs n'ont pas été intégrés. Leur introduction agrandissait démesurément les groupes de liaison II, IV, VI et IX ($d > 100$ cM). Généralement, cela est la conséquence d'erreurs de génotypage, de données manquantes mais aussi de la stratégie de génotypage sélectif qui peuvent créer de fausses liaisons entre marqueurs (Lod score > 3 ; Lin et al., 1996). L'échantillonnage, parmi les parties extrêmes de la courbe de distribution, n'est donc pas représentatif de la population dans son ensemble notamment au niveau des fréquences alléliques de marqueurs. De telles distances génétiques ne peuvent refléter la réalité car la distance maximum entre deux marqueurs ne doit pas dépasser un taux de recombinaison supérieur à 0,5.

4.5.2 Comparaison de cartes au sein de l'espèce *Crassostrea gigas*

Comme précédemment cité, deux cartes de liaison ont été établies chez l'huître creuse *Crassostrea gigas* (Hubert et Hedgecock, 2004 ; Li et Guo, 2004). La taille et la couverture du génome de la carte consensus développée ici sont en accord avec ces précédentes cartes avec une couverture du génome estimée à environ 80%. Notre carte consensus comporte un nombre de groupes de liaison égal au nombre de chromosomes ($n=10$) chez l'espèce, contrairement aux cartographies publiées précédemment où 11 et 12 groupes de liaison ont été respectivement formés (Hubert et Hedgecock, 2004 ; Li et Guo, 2004). De plus, la densité et le nombre des marqueurs de notre carte consensus sont légèrement inférieurs, avec un nombre total de 80 marqueurs cartographiés et un espace inter-marqueur de 13,3cM en moyenne. En effet, nous avons choisi au préalable des marqueurs microsatellites répartis de manière relativement régulière sur la carte de liaison publiée par Hubert et Hedgecock (2004). De plus, nous avons également rencontré des difficultés d'amplification par PCR de certains marqueurs microsatellites (voir chapitre III). Cela explique également qu'il n'existe pas de corrélation entre le nombre de marqueurs cartographiés et la longueur des groupes de liaison entre notre carte consensus et la carte publiée par Hubert et Hedgecock (2004). Cependant, l'organisation générale des deux cartes est relativement proche car il existe une corrélation entre la distance génétique moyenne entre les paires de marqueurs communes.

L'intérêt de nos travaux se situe tout d'abord dans la confirmation de la carte développée à partir de marqueurs microsatellites par Hubert et Hedgecock (2004) mais surtout dans l'intégration de marqueurs de type I dans cette cartographie génétique. A ce jour et à notre connaissance, aucun marqueur de ce type n'a été cartographié chez *C. gigas*. Ces marqueurs présentent donc un réel intérêt puisqu'ils reflètent, en théorie, l'emplacement de gènes dans le génome de l'espèce d'autant plus si un marqueur de ce type se situe dans une

région chromosomique liée à la variance d'un trait (région QTL). Les marqueurs de type I sont des marqueurs relativement transférables entre laboratoires mais aussi entre espèces relativement proches telles que l'huître creuse *C. virginica* ou l'huître plate européenne *Ostrea edulis*. En effet, l'utilisation potentielle d'amorces, initialement développée chez *C. gigas*, pour la détection de SNP par amplification croisée chez *O. edulis* a été décrite par Lallias et collaborateurs (2006).

En général, les cartes génétiques de densité moyenne à faible s'étendent sur une longueur supérieure à celle de cartes de densité importante. Cette observation est particulièrement frappante chez le medaka, *Oryzias latipes*, où une carte composée de 170 marqueurs s'étend sur 2480 cM (Wada et al., 1995) alors que l'utilisation de 663 marqueurs aboutit à la construction d'une carte de 1354 cM (Naruse et al., 2000). L'effet de la densité moyenne des marqueurs de notre carte consensus tend donc à augmenter la taille des groupes de liaison. En effet, notre carte consensus, qui comporte 80 marqueurs, s'étend sur une longueur de 1062 cM ce qui est supérieur aux cartes précédemment publiées qui comportaient respectivement 119 et 102 marqueurs pour des distances respectives de 1030 et 770 cM (Hubert et Hedgecock, 2004; Li et Guo, 2004). De plus, les erreurs de génotypage tendent également à augmenter la taille des groupes de liaison (Collins et al., 1996). L'ajout de marqueurs supplémentaires, de type SNP ou microsatellite, permettra très certainement d'augmenter la densité de cette carte de liaison consensus mais aussi de réduire sa taille et de la consolider pour en faire une solide base de travail.

4.5.3 Recombinaison différentielle entre mâles et femelles

Les premières études des différences de taux de recombinaison entre sexes ont été menées chez la drosophile, il y a pratiquement un siècle (Morgan, 1914). En général, cette différence n'est pas constante le long du génome (Lagercrantz et al., 1995). Les régions entourant le centromère montrent un taux de recombinaison plus élevé chez la femelle (Broman et al., 1998; Mohrenweiser et al., 1998) tandis que chez le mâle, ce sont les régions télomériques (Zahn et al., 1995; Broman et al., 1998). Plusieurs hypothèses ont été avancées afin d'expliquer ce phénomène. Cependant, peu d'entre elles ont aboutit à une conclusion claire. Une pression de sélection en phase gamétique, à l'origine des différences de recombinaison entre sexe, est l'hypothèse la plus vraisemblable (Korol et al., 1994). Cette hypothèse est supportée par de nombreux auteurs (Singer et al., 2002; Lenormand et al., 2005; Lorch, 2005). Plus concrètement, l'évolution adaptative tend à réduire le taux de recombinaison chez le mâle car la compétition pour se reproduire s'intensifie au sein des

individus mâles de l'espèce. Dans ce cas, seuls les mâles dont la combinaison allélique est favorable à la fitness de ses descendants ont plus de chance de transmettre cette combinaison allélique et de donner une descendance viable. Un taux de recombinaison trop important tend à séparer ou détruire cette combinaison favorable d'allèles. Une balance s'établit entre une situation où la recombinaison est faible et une seconde situation où la recombinaison est trop intense.

Cependant, l'huître creuse est un animal hermaphrodite à tendance protandrique. Le nombre de gamètes produits par le mâle est en moyenne supérieur d'un facteur 10 à celui des femelles. De plus, le génome de cette espèce ne comporte pas de chromosome sexuel et peu de connaissances ont été acquises sur le déterminisme génétique du sexe (Guo et al., 1998). De part la forte fécondité des mâles, on peut penser que la recombinaison est plus importante chez le mâle que chez les femelles. Celles-ci produisent moins de gamètes, d'où un nombre d'événements de recombinaison moins important. Or l'ensemble des cartographies génétiques établies chez les mollusques bivalves, et plus généralement dans le monde vivant, met en évidence un taux de recombinaison plus important chez la femelle (Yu et al., 2003; Hubert et Hedgecock, 2004; Li et Guo, 2004; Lallias et al., 2007). Le ratio de recombinaison observé chez Hubert et Hedgecock est légèrement inférieur à celui observé sur la carte consensus avec un ratio entre la longueur totale de la carte mâle vs femelle, de 1,26 et 1,36 respectivement. Cependant, les mécanismes moléculaires responsables des différences de taux de recombinaison entre les deux sexes sont mal connus chez l'huître creuse (Hubert et al., 2004). Enfin, la différence de taux de recombinaison entre les genres peut être exploitée afin d'améliorer la localisation de régions QTL et de différencier deux QTL situés à la proximité l'un de l'autre car la résolution des cartes sexe spécifiques sont différentes (Lorch, 2005).

4.6 Perspectives de cartographie génétique chez *Crassostrea gigas*

4.6.1 Poursuite du développement de marqueurs moléculaires

L'un des objectifs des travaux de cartographie de liaison est de situer les régions chromosomiques en relation avec la variance d'un trait phénotype (région QTL). Plus cette localisation est précise, plus le nombre de gènes compris dans un intervalle sera petit et par conséquent, plus la recherche des gènes responsables de cette variation sera aisée. Dans cet objectif, la carte de liaison la plus adéquate doit contenir un grand nombre (dépendant de la taille du génome) de marqueurs répartis assez régulièrement au sein de cette carte. En s'appuyant sur les ressources en développement (banques EST) chez *C. gigas*, les prochaines

étapes vers l'établissement d'une carte de liaison dite saturée consistent à l'identification de marqueurs SNP et microsatellites (voir chapitre III).

A moyen terme, une carte génétique peut également servir de base à l'étude de la synténie (conservation de l'ordre des gènes le long des chromosomes) au sein du genre *Crassostrea* et définir des zones de synténie ou des zones « floues » (sans correspondance entre les cartes). A plus long terme, lorsque qu'un panel de cartes génétiques sera disponible chez différentes espèces proches de l'huître creuse, des approches de cartographie et de génomique comparée pourront être envisagées afin d'étudier analyse comparative de l'organisation chromosomique des génomes, de comprendre le fonctionnement des structures biologiques et d'en inférer l'évolution pour mettre en évidence les remaniements structuraux subis par le génome des espèces. Cependant, le coût expérimental actuel de l'analyse d'un génome restant très élevé, les informations disponibles sont le plus souvent concentrées sur quelques organismes « modèles » eucaryotes.

4.6.2 Intégration des différents types de cartes

Au-delà des cartes de liaison, très peu de travaux ont été entrepris dans la construction de carte physique, cytogénétique ou d'hybrides de radiation chez les bivalves. Cependant, chez les poissons tels que le saumon atlantique (Thorsen et al., 2005) ; le catfish (Quiniou et al., 2003) ; la carpe (Katagiri et al., 2001) ; le tilapia (Katagiri et al., 2005) ; la truite arc-en-ciel (Katagiri et al., 2001; Phillips et al., 2003; Palti et al., 2004), le bar (Whitaker, 2006) ou encore chez le medaka (Matsuda et al., 2001) des banques BAC sont disponibles. Chez le zebrafish (Phillips et al., 2001) ; la truite arc-en-ciel (Phillips et al., 2005) et le saumon atlantique (Mitchell et al, unpublished), la localisation cytogénétique de gènes a été effectuée. Ces travaux constituent d'intéressantes pistes à suivre pour l'établissement de telles cartes chez *C. gigas*. Lorsqu'un plus grand nombre de marqueurs moléculaires sera disponible, cela facilitera les travaux de cartographies cytogénétiques. Enfin, une carte physique d'assemblage de BACs (« fingerprinting ») est en cours de construction chez *C. gigas* (Gaffney, pers. com.).

Chapitre V

CARTOGRAPHIE DE

CARACTERES A

DETERMINISME

MULTIGENIQUE

Les données de phénotypage de deux caractères (chapitre II), de génotypage (chapitre III) et la cartographie génétique consensus (Chapitre IV) ont été combinées pour rechercher des QTL liés à la survie et à l'infection à l'herpès virus en période estivale. Les méthodes et résultats sont décrits dans ce chapitre V.

5.1 Principe de la détection de QTL

L'utilisation de marqueurs génétiques pour identifier les gènes affectant des caractères biologiques complexes est appelée analyse QTL (Lynch et Walsh, 1998). Un QTL (**Quantitative Trait Locus**) se définit au sens large comme un segment chromosomique contenant un ou plusieurs gènes influençant un phénotype ou caractère dans une certaine proportion ou variance. Cette variance génétique peut être décomposée en variance additive (A), variance de dominance (D) et l'interaction (I) de ces deux éléments ($V_G = V_A + V_D + V_I$). Ces éléments peuvent se définir comme tel : l'effet moyen d'un allèle est l'espérance centrée de la valeur phénotypique des individus ayant reçu cet allèle d'un de leurs parents, l'autre allèle étant tiré au hasard. Cet effet moyen est également désigné comme l'effet additif d'un allèle. La valeur génétique additive d'un individu est la somme des effets moyens des gènes qu'il possède. L'effet moyen d'un allèle particulier est l'espérance de la valeur génétique conditionnée par la présence de cet allèle dans le génotype. Cette valeur minimise l'espérance du carré du résidu de dominance. La valeur génétique additive (A) représente la fraction de la valeur génétique (G) dont il est possible de prédire la transmission de parent à descendant : un parent transmet, en espérance, la moitié de sa valeur génétique additive.

La cartographie de QTL est l'association entre un ou des génotype(s) à des localisations génomiques différentes et un phénotype pour un ensemble de caractères quantitatifs en terme de nombre, de position, d'effets et d'interactions. La distribution, le nombre, l'effet et la position physique au sein du génome des gènes affectant le caractère considéré définissent la notion d'architecture génétique d'un caractère. Le mode d'action des gènes et plus particulièrement si leurs effets sont constant, cet effet est dit additif. Si l'effet est contingent sur l'interaction avec d'autres gènes, il s'agit d'épistasie ou si le gène induit plusieurs effets phénotypiques apparemment indépendants, on parle de pléiotropie (Erickson, 2005).

Un QTL ne peut souvent se limiter à un unique locus du fait de la faible résolution de la plupart des cartes de liaison actuellement disponibles. Un QTL représente donc le plus souvent un segment de chromosome qui peut contenir un ou plusieurs loci affectant le caractère phénotypique considéré. De plus, il a été démontré dans de nombreuses études revues par Falconer et Mackay (1996) que les QTLs ségrégent selon les lois mendéliennes et que la majorité des caractères génétiques sont polygéniques. Le phénotype observé, pour de tels caractères, est le résultat de complexes interactions entre des allèles multiples et de facteurs environnementaux. L'intérêt final des analyses QTL est donc d'identifier le/les gène(s) qui contrôle ce caractère. L'analyse des QTLs peut être divisée en deux étapes : (1) du phénotype au génotype(s) et QTL ; (2) du QTL au(x) gène(s). La première étape consiste à corrélérer les marqueurs génétiques à un phénotype dans une population ségrégeante. Ensuite, le but est de réduire la région corrélée au phénotype à un intervalle aussi petit que possible ne contenant alors que quelques gènes. Ce processus de réduction de la région chromosomique est rarement couronné de succès car il est très difficile de réduire cette région à un intervalle contenant quelques gènes, voir, un unique gène (Darvasi, 1998) mais devient techniquement de plus en plus facilement réalisable.

5.2 Etapes des analyses QTL

La première étape, décrite ci-dessus, peut être subdivisée en 3 étapes : la détection de QTL, la cartographie de QTL et la cartographie fine de QTL.

- (1) la détection de QTL : Cette détection correspond à l'étude de la part de variance phénotypique expliquée par le QTL. Un paramètre tel que la taille varie au sein d'une même population et montre une variance. Cette variance est la conséquence d'une part, de l'environnement et d'autre part, de la variation génétique. La contribution d'un unique allèle peut être mesurée comme la proportion de variance qu'il explique. Un QTL qui explique une large part de variance phénotypique sera plus facilement identifiable. Le cas extrême serait un gène qui explique à lui seul toute la variance phénotypique. Il s'agirait alors d'un gène majeur.
- (2) la seconde étape consiste à la localisation chromosomique des régions significativement affectées par la variation du caractère dans la population. Cette localisation est importante pour l'identification des gènes responsables de la variation et la compréhension des mécanismes génétiques sous jacents à cette variance.

- (3) L'identification des gènes causaux et les mutations causales ou QTN (Quantitative Trait Nucleotide).

Peu de caractères phénotypiques ont été analysés depuis la localisation des QTLs jusqu'à l'identification de gènes ou de mutations causales. Cet objectif a été atteint chez la tomate (Fridman et al., 2004) pour la teneur en sucre ou la souris pour la résistance à des tumeurs intestinales (Cormier et al., 1997). Des paramètres tels que la taille de la population ou l'effet du QTL influencent sa détection (Darvasi, 1998). La détection de QTL dans de petites populations est une difficulté couramment rencontrée chez les animaux (Holland, 2007) car l'effet du QTL est confondu avec les artéfacts statistique liés à l'échantillonnage. Chez la drosophile, l'architecture génétique simple de la régulation de la Dopa decarboxylase, mise en évidence à partir d'un petit échantillon d'individus, a été totalement revue par la cartographie fine de QTLs sur un échantillon beaucoup plus important (Mackay, 2004). Chez le maïs, la recherche de QTL liés au rendement du grain a démontré que l'architecture génétique de ce caractère correspond à un grand nombre de QTLs d'effets faibles et qui n'expliquent pas plus de 50% de la variance phénotypique au total. Cela implique que ce caractère est sous l'influence d'un grand nombre de gènes ce qui rend sa cartographie et l'étude de son architecture très complexe (Schon et al., 2004). La recherche de QTL n'explique jamais l'ensemble de la variance phénotypique d'un caractère, sauf s'il s'agit d'un gène majeur comme précédemment décrit car les modèles de génétiques quantitatives ne parviennent pas à mettre en évidence un grand nombre de QTL à effets faibles, ceux-ci ne parvenant tout simplement pas à les détecter (Holland, 2007).

Le domaine de recherche des QTLs est relativement vaste : il s'étant de l'analyse d'un caractère en agriculture tel que la croissance ou la fitness, à l'étude de la prédisposition au cancer (Falconer et al., 1996) en passant par la résistance au froid (Perry et al., 2001) ou l'intelligence (Plomin et al., 1994) et s'applique à un très grand nombre d'espèces.

5.3 Modèles de cartographie de QTL

Sax (1923) fut le premier à utiliser des marqueurs de pigments chez le haricot afin d'analyser les gènes affectant la taille des graines par le biais des taux de recombinaison chez des descendants F2 issus de différents croisements. Puis, Thoday (1961) proposa d'utiliser deux marqueurs bordant de part et d'autre le locus étudié afin de le localiser. L'idée de Sax et Thoday reposait donc sur l'association entre le QTL ou segment chromosomique et un

ensemble de marqueurs. Le développement de cartographies génétiques établies chez divers organismes a grandement facilité la cartographie et l'analyse de QTL(s). Plus récemment, Lander et Botstein (1989) ont proposé une méthode plus robuste pour la cartographie des QTLs : « l'**Interval Mapping** » (IM). Cette méthode est basée sur l'utilisation d'un couple de marqueurs. Au sein de l'intervalle défini par les deux marqueurs choisis, la position du QTL est testée pour chaque position dans cet intervalle. L'approche d'« Interval Mapping » détecte un QTL à la fois lors de l'étape de cartographie. Si la carte de liaison disponible a une résolution relativement importante, un « scan » complet du génome peut être réalisé, afin de produire un profil statistique. La position la plus vraisemblable du QTL est alors déduite statistiquement. De plus, il a été démontré que cette méthode d'« interval mapping » est relativement robuste et nécessite peu de descendants (Lander et Botstein, 1989; Zeng, 1994). Enfin, les modèles basés sur la régression multiple donnent des résultats très semblables à l'IM (Haley et Knott, 1992). L'utilisation de la méthode de régression pour la détection de QTL, présente plusieurs avantages par rapport à l'utilisation des méthodes basées sur le maximum de vraisemblance (Maximum Likelihood). La méthode de régression nécessite moins de temps de calcul, permet l'implantation de variables en tant qu'effets fixes ou cofacteurs, et permet la détection de QTL dans des schémas relativement complexes (Knott, 2005).

L'avantage de la méthode d'IM est de localiser sur une carte génétique, de façon précise, la position la plus vraisemblable d'un QTL. La densité des marqueurs n'est le facteur principal qui influence la détection des QTL. Darvasi et collaborateurs (Darvasi et al., 1993) ont montré que la puissance de détection et la précision de la localisation du QTL est directement influencé par la stratégie d'échantillonnage. Ceci a été vérifié expérimentalement par Song et collaborateurs (Song et al., 1995). Ainsi, si la couverture de l'ensemble du génome est relativement moyenne (un marqueur tous les 25 cM), il est plus judicieux d'augmenter le nombre d'individus que de réduire l'espace inter-marqueurs. Les deux principales limites du modèle "Interval Mapping" sont, d'une part la sensibilité aux distributions non normales des caractères quantitatifs et, d'autre part qu'il reste un modèle ne proposant la détection que d'un seul QTL. En effet, lorsque deux QTL ségrégent l'un au voisinage de l'autre, et lorsqu'ils agissent dans la même direction, un seul QTL fantôme peut être détecté au milieu des deux QTL (Martinez et Curnow, 1992). Lander et Botstein (1989) suggèrent que la présence de deux QTL assez éloignés (50 à 80 cM) sur le même groupe de liaison peut être distinguée empiriquement en fixant la position d'un QTL et en recalculant le

LOD score pour l'autre et vice versa. Une augmentation des LOD de deux unités indiquerait la présence de deux QTL distincts. D'autre part, Martinez et Curnow (1992) et Zeng et collaborateurs (1993) ont montré qu'il était possible de détecter des QTL proches par régression multiple.

Cependant, cette méthode peut biaiser l'identification et la position du QTL lorsque de multiples QTLs se situent dans le même groupe de liaison (Haley et Knott, 1992; Zeng, 1994). Afin de résoudre ce problème, Jansen (1993) et Zeng (1993) ont indépendamment proposé l'idée de combiner la méthode d'intervalle mapping avec une analyse en régression multiple afin d'obtenir une analyse dite en « **Composite Interval Mapping** » ou CIM. Lors de ce test, d'autres marqueurs sont utilisés en tant que co-variants afin de réduire la variance résiduelle de QTLs situés à proximité et ainsi améliorer le test. Ce modèle utilise l'association d'un QTL et de plusieurs marqueurs. Enfin, Kao et collaborateurs (1999) ont développé une nouvelle méthode nommée « **Multiple Interval Mapping** » (MIM) toujours basée sur l'« Interval Mapping ». Cette méthode utilise simultanément des intervalles entre des couples de marqueurs afin de détecter de multiples QTLs. Selon les auteurs, la MIM est plus performante et plus précise dans la détection des QTLs que les autres méthodes. En outre, elle est capable de mettre en évidence des effets tels que l'épistasie et d'estimer l'héritabilité des caractères considérés. De plus, la variance génétique contribuant à chaque QTL est également estimée permettant ainsi d'orienter une sélection assistée par marqueurs. A l'instar des logiciels de cartographie génétique, un grand nombre de logiciels de recherche de QTL ont été créés. Manly et collaborateurs (1999) ont établi un inventaire des principaux logiciels et des méthodes de détection utilisées (Tableau 16).

5.4 Paramètres influençant les analyses QTL

Divers facteurs influencent la puissance de détection des QTLs afin d'identifier les loci à la base de la variance des caractères phénotypiques. Ces facteurs incluent le schéma expérimental, le type de marqueurs moléculaires et la taille d'échantillons.

5.4.1 Nombre et type de marqueurs

Il existe deux types de marqueurs moléculaires couramment utilisés dans les analyses QTL : les dominants et les co-dominants. Le type de marqueurs moléculaires affecte la résolution des QTLs car les marqueurs co-dominants ont une puissance estimée à 20 fois supérieure aux marqueurs dominants (Liu, 1998).

Tableau 16 : Principaux logiciels de détection de QTL (Manly et al., 1999)

Logiciel	Fonctions	Méthodes	Designs	Publications associées	Site internet
Mapmaker/QTL	SIM, CIM	ML	BC, F2	Lincoln et al. (1992)	ftp://genome.wi.mit.edu/pub/mapmaker3
QTL Cartographer	SIM, CIM	ML	diverse	Basten et al. (1997)	http://statgen.ncsu.edu/qtlcart/
Map Manager QT	SIM, CIM	LS	BC, F2	Manly and Olson (1999)	http://mcbio.med.buffalo.edu/mapmgr.html
MapQTL ® 5	SIM, CIM	ML	BC, F2, DH	Van Ooijen (2004)	http://www.cpro.dlo.nl/cbw/
PLABQTL	SIM, CIM	LS	BC, F2	Utz and Melchinger (1996)	http://www.uni-hohenheim.de/~ipspwww/soft.html
MQTL	SIM, CIM	LS	BC, DH	Tinker and Mather (1995)	ftp://gnome.agrenv.mcgill.ca/pub/genetics/software/MQTL
Multimapper	CIM	Bayesian	BC, F2	Sillanpaa and Arjas (1998)	http://www.rni.helsinki.fi/~mjs
QTL Express	SIM	LS	Inbred, Outbred	Seaton et al. (2002)	http://qtl.cap.ed.ac.uk

SIM : Single Interval Mapping ; CIM : Composant Interval Mapping ; ML : Maximum Likelihood ; LS : Least Square ; BC : Backcross ; F2 : Famille de 2^{nde} génération ; DH : Double Hybride

Les marqueurs lus comme dominants tels que les AFLPs vont produire deux classes génotypiques, au lieu de trois, du fait de la dominance, de telle sorte qu'ils ne permettent pas de distinguer entre un hétérozygote et un homozygote dominant. Il y a donc une perte d'information liée au nombre d'évènements de recombinaison observés qui se restreint dans l'intervalle considéré lorsqu'un tel type de marqueur est utilisé. Cependant, l'utilisation des marqueurs AFLP est plus simple car elle s'affranchit du séquençage et du développement de nouveaux marqueurs, tout en permettant de construire des cartes de liaison couvrant en grande partie le génome. De plus, au cours d'une même réaction de PCR, de nombreux loci peuvent être amplifiés simultanément multipliant ainsi le nombre de marqueurs. Cependant, nous n'avons pas choisi d'utiliser ce type de marqueur moléculaire étant donné leur nature dominante.

Les marqueurs co-dominants, tels que les microsatellites ou les SNP offrent une plus grande puissance dans la détection de la recombinaison entre deux marqueurs adjacents (Liu, 1998). Cependant, leur coût supérieur de développement est contre balancé par cette puissance supérieure amenant à une détection plus fine des QTL tant en terme de variance, que de position. Les microsatellites restent le marqueur de choix pour les analyses QTL car ils ont été développés pour un très grand nombre d'espèces d'intérêt. Les SNP sont plus couramment utilisés chez des organismes très étudiés comme l'homme ou le poulet pour lesquels 5 et 1,42 millions de SNP sont disponibles à ce jour respectivement (Sachidanandam et al., 2001; Ka-Shu Wong et al., 2004).

5.4.2 Distribution des marqueurs

De même que le type de marqueurs moléculaires, la distribution de ceux-ci affecte la résolution et l'effet des QTLs. Plus le nombre de marqueurs disponibles est important, plus précise sera l'estimation de l'effet et de la position du QTL. Cependant, un juste milieu entre le nombre de marqueurs et la taille de l'intervalle moyen qui les sépare doit être trouvé. Dans la plupart des études, les conditions optimales décrites correspondent à un intervalle de 10 à 15 cM et un échantillonnage de 300 individus (Lynch et Walsh, 1998). En effet, si la distance intermarqueur est inférieure à quelques centimorgans (2 à 5 cM) le taux de recombinaison moyen entre paires de marqueurs déclinera affectant les analyses. Dès lors, des génotypes dits « non répliqués » peuvent apparaître et poser problème (Knapp et al., 1990).

5.4.3 Echantillonnage et schéma expérimental

En général, les analyses QTL employant des schémas expérimentaux classiques et un nombre important d'échantillons ($N > 300$) permettent d'identifier des QTL dit « à effet large », c'est à dire où, au moins, 15% de la variance totale du caractère est expliquée par le QTL (Liu, 1998; Lynch et Walsh, 1998; Doerge, 2002). En théorie, un QTL dont l'effet additif représente 5% de la variance peut être détecté à partir d'un échantillon de 206 individus issus d'une F2 génotypés par des marqueurs co-dominants si ceux-ci sont distants de 5cM. Cependant, en raison d'une faible héritabilité du caractère considéré et des génotypes et phénotypes incomplets, le nombre estimé d'individus nécessaires à la détection de QTL atteint 300 (Doerge et al., 1997). Ensuite, l'estimation de l'effet du QTL peut être biaisée par les méthodes statistiques employées : c'est l'effet Beavis (1998). En effet, lorsque l'échantillonnage est inférieur à 100 individus, l'estimation de l'effet du QTL est surévaluée et, à l'inverse, la puissance de détection de QTL à effets faibles diminue rapidement. Ce biais statistique tend à identifier de faux QTLs et à surestimer l'effet des véritables.

Tout comme le nombre d'individus, le type de croisement affecte également la détection des QTLs. Dans l'exemple décrit ci dessus, si le croisement était un backcross au lieu d'une descendance F2, le nombre d'individus nécessaire serait deux fois supérieur afin d'obtenir la même précision dans la détection du QTL (Lynch et Walsh, 1998). Parmi les différents schémas expérimentaux de croisement, les lignées hybrides (ou inbred lines) sont plus adaptées à la détection de QTL car cette méthode tend à augmenter le déséquilibre de liaison entre les marqueurs et le QTL (Doerge et al., 1997; Liu, 1998; Mackay, 2001). Ce schéma nécessite deux individus ou deux familles d'individus présentant un contraste pour le phénotype étudié mais aussi pour les marqueurs utilisés. Le croisement entre ces deux individus permet de générer une famille F1 qui peut être recroisée afin de donner une génération F2 dite « recombinante ».

5.4.4 Le génotypage sélectif

Certains descendants sont plus informatifs que d'autres : de façon générale, les individus les plus informatifs sont ceux dont le génotype est relié sans ambiguïté à leur phénotype. Par exemple, Lander et Botstein (1989) ont démontré que la majorité de l'information de liaison concernant les maladies humaines provenait des individus atteints par ces maladies contrairement aux porteurs sains. Etant donné que le phénotype des individus non affectés n'est pas certain, ceux-ci n'apportent que peu d'information. En remplaçant cette

constatation dans le contexte de la génétique quantitative, l'information provient donc essentiellement des individus dont le phénotype est caractérisé comme « extrême » c'est à dire celui s'écartant le plus du phénotype classique ou moyen. Il est donc plus judicieux de génotyper les individus dont les phénotypes sont considérés comme extrême.

Cette stratégie présente l'avantage de réduire considérablement les coûts du génotypage. En effet, les individus ayant des performances supérieures de un écart-type par rapport à la moyenne, c'est-à-dire 33% des individus, contribuent à 81% de l'information totale (Lander et Botstein, 1989). Cependant, cette méthode présente des inconvénients majeurs. Premièrement une sélection biaisée des individus extrêmes entraîne une sur-estimation de la variance phénotypique expliquée par le ou les QTL (Lander et Botstein, 1989). Deuxièmement, la caractérisation sélective rend difficile l'estimation de l'effet de substitution allélique au locus du QTL. Cependant, en tenant compte de la proportion de la population échantillonnée dans les queues de la distribution et de la différence de performance observée entre elles, Darvasi et Soller (1992) ont montré qu'il était possible d'estimer l'effet phénotypique du QTL.

5.4.5 Seuil de signification

Dans le cadre des modèles basé sur l'« Interval Mapping », des seuils de signification relativement élevés ($\alpha=0.05$ ou $\alpha =0.01$) ont généralement été utilisés (Soller et al., 1976; Edwards et al., 1987; Reiter et al., 1992). Si ces seuils apparaissent suffisants pour un test isolé, ils ne le sont pas pour des marqueurs liés sur un groupe de liaison. Un niveau de signification plus strict est en effet nécessaire, pour éviter d'accepter un grand nombre de faux QTL. Il en résultera par contre une puissance de détection (probabilité d'accepter un vrai QTL) diminuée. Cependant, le choix d'un risque assez faible est essentiel si l'objectif de l'expérience de détection de QTL vise à identifier ou introgresser des gènes. De même, lorsque l'« Interval Mapping » est appliqué à l'ensemble des marqueurs liés sur une carte génétique, pour assurer un risque global de 5% pour l'ensemble du génome, il convient d'utiliser un risque beaucoup plus faible au niveau de chaque intervalle pour tenir compte de la non indépendance des tests pour les marqueurs appartenant au même groupe de liaison. Lander et Botstein, (1989) ont proposé une approximation du seuil, au niveau du génome, basée sur sa taille (en cM) et sur le nombre de chromosomes de l'espèce étudiée. Rebaï et al (1994) ont développé une méthode permettant de calculer analytiquement un seuil approximatif pour chaque groupe de liaison en tenant compte de leur longueur et de leur

densité en marqueurs. Par exemple, pour un seul chromosome de 100 cM et 200 individus génotypés, le seuil de ce test à 5 % est de 6,90 pour 5 marqueurs equirépartis et 9,05 pour une infinité de marqueurs. Dans une échelle lod-score, ces seuils sont de 1,5 et 1,97. Notons que certains auteurs ou logiciels préconisent l'usage d'un seuil de 3 dans l'échelle lod-score ce qui correspond donc à un seuil plus sévère pour chaque chromosome. Enfin, Churchill et Doerge (1994) ont quant à eux proposé de déterminer ce seuil par des tests de permutation.

5.4.6 Estimation de l'intervalle de confiance de la localisation des QTL par « Bootstrapping »

Toutes les méthodes de cartographies de QTL sont basées soit sur un maximum de vraisemblance (Lander et Botstein, 1989), soit sur une méthode de régression (Haley et Knott, 1992; Martinez et Cudnoz, 1992). Cependant, ces méthodes n'intègrent pas le calcul de l'intervalle de confiance de la position du QTL. Visscher et collaborateurs (1996) proposent d'estimer cet intervalle par une méthode de bootstrapping. Le bootstrapping consiste au remplacement d'échantillons (ici association marqueur-génotype) dans la distribution. A chaque cycle de calcul, N échantillons sont intervertis et la position du QTL est recalculée. Cette méthode de bootstrapping est assez robuste si le nombre de cycles de calcul est supérieur à 1000. De plus, elle est applicable aux méthodes de maximum de vraisemblance et de régression car les résultats obtenus sont très similaires mais aussi pour la recherche de QTLs multiples (Visscher et al., 1996).

5.4.7 Part de variance expliquée

Généralement, les résultats d'une recherche de QTL se résument à indiquer le groupe de liaison dans lequel le QTL a été mis en évidence, sa position et son intervalle de confiance (exprimée en cM) mais aussi la part de variance qu'il explique (exprimée en %). Cette part de variance qui reflète l'importance du QTL se calcule à partir des sommes des carrés (Residual Sum Square - RSS) du modèle réduit et du modèle complet selon la formule suivante : $4 \cdot (1 - \text{RSS Mod\`e le complet} / \text{RSS Mod\`e le r\`e duit})$.

Tableau 17 : Revue des QTL détectés chez les principales espèces d'intérêt aquacole

Espèce	Caractère(s)	Nombre de QTL détectés	Part de variance (%)	Références bibliographiques
Truite arc en ciel - <i>Oncorhynchus mykiss</i>	Tolérance à la température	2	22	(Jackson et al., 1998)
	Tolérance à la température	3		(Danzmann et al., 1999)
	Résistance au virus INHV	2		(Palti et al., 1999)
	Temps de ponte	5		(Sakamoto et al., 1999)
		2	24,6	
	Taille embryonnaire	2	22,6	(Robison et al., 2001)
	Poids embryonnaire	2	26,2	
	Résistance virale (IPNV)	2	27-34	(Ozaki et al., 2007)
	Temps de ponte	4	août-64	(O'Malley et al., 2003)
	Poids	3	déc-25	
	Poids	5	04-juil	(Martyniuk et al., 2003)
	Résistance virale (IHNV)	3		(Rodriguez et al., 2004)
	Taux de croissance embryonnaire	7	0,6-23	(Nichols et al., 2007)
	Résistance virale (IHNV)	1		(Khoo et al., 2004)
	Nombre de pilores	3	13-19	(Zimmerman et al., 2005)
	Niveau de cortisol	2	43	(Drew et al., 2007)
	Poids	-	-	(Fotherby et al., 2007)
	Age de maturation sexuelle	-	-	
	Maturation	-	-	(Haidle et al., 2007)
	Saumon Atlantique - <i>Salmo solar</i>	Résistance virale	10	27,3 (total)
Résistance virale (ISA)		2		(Moen et al., 2007)
Poids		2	20,1	(Reid et al., 2005)
Conditionnement		4	24,9	
Pourcentage de lipide				(Derayat et al., 2007)
	Résistance virale (IPN)	2		(Houston et al., 2008)
Omble chevalier - <i>Salvelinus alpinus</i>	Maturation sexuelle	3	23	
	Poids	3	34	(Mogahadam et al., 2007)
	Conditionnement	1		
Bar - <i>Dicentrarchus labrax</i>	Traits morphométriques	1	-	(Chatziplis et al. 2007)
Carpe commune - <i>Cyprinus carpio</i>	Tolérance au froid			(Sun et al., 2004)
Huitre creuse - <i>Crassostrea gigas</i>	Heterosis			(Hedgecock et al., 2007)
Huitre plate - <i>Ostrea edulis</i>	Résistance à la bonamiose	3		(Lallias, 2006)
Ormeau - <i>Haliotis rubra</i>	Taux de croissance			(Baranski et al., 2007)
Pétoncle de baie - <i>Argopecten irradians</i>	Taille			(Qin et al., 2007)

5.5 Recherche de QTLs chez les espèces d'intérêt aquacole

Dans la littérature relative à la recherche de QTL chez les espèces d'intérêt aquacole, une importante partie de ces travaux ont été menés chez la truite arc en ciel (*Oncorhynchus mykiss*) pour laquelle des caractères d'intérêts, tels que la période de ponte, le temps de maturation, la résistance à divers pathogènes, la croissance ou la tolérance à des stress ont été étudiés. Néanmoins, la diversité des espèces impliquées dans des programmes de recherche de QTL est en constante augmentation. Récemment, des travaux ont été menés chez le bar, *Dicentrarchus labrax* (Chatziplis et al., 2007), l'ormeau, *Haliotis rubra* (Baranski et al., 2007) ou l'huître creuse *Crassostrea gigas* (Hedgecock et al., 2007). Le tableau 17 dresse une revue récente des travaux menés en recherche de QTLs chez les principales espèces d'intérêt aquacole.

La plupart des études localisent les régions QTLs dans un intervalle de 10 à 60 cM. Cet intervalle est bien trop important pour permettre une localisation fine des gènes sous jacents à la variation du/des caractère(s) phénotypique(s) considéré(s). C'est pourquoi, la réduction de cet intervalle est une étape importante et nécessaire vers l'identification de ces gènes et à plus long terme pour une application en sélection assistée par marqueurs. La sélection assistée par marqueur est une stratégie pleine de promesses qui doit être appliquée chez une espèce d'intérêt aquacole. La résistance à un syndrome ou un pathogène est un caractère intéressant pour le développement d'une stratégie de sélection assistée par marqueurs. Par exemple, chez la truite arc en ciel, plusieurs études ont mis en évidence des régions QTLs liées à la résistance à un pathogène (Palti et al., 1999 ; Khoo et al., 2004 ; Rodriguez, 2004 ; Ozaki et al., 2007).

Le succès d'une telle stratégie repose principalement sur l'héritabilité du caractère et la proportion de variance additive expliquée par les marqueurs. Les résultats obtenus jusqu'à présent en recherche de QTL chez les poissons démontrent l'intérêt de rechercher des QTL chez un panel d'espèces plus large et plus particulièrement les mollusques bivalves de part leur intérêt commercial. Sakamoto et collaborateurs (1999) confirment que l'identification de marqueurs moléculaires liés à la variance de caractères d'intérêt économique tend à développer des schémas de sélection assistée par marqueurs. Cependant, le succès de ce schéma de sélection repose sur la connaissance de l'architecture génétique des caractères étudiés. Des interactions épistatiques tendent à limiter l'application des schémas de sélection assistée par marqueurs (Danzmann et al., 1999). Il apparaît donc important d'identifier les

interactions entre les gènes sous jacents à la variation des caractères phénotypiques suite à l'identification de QTL(s) afin d'engager, ou non, un schéma de sélection assistée par marqueurs.

5.6 Matériel et méthodes

5.6.1 Simulations de la puissance de détection des QTL

L'influence de tous les paramètres des analyses QTL a été testée par le biais de simulations réalisées par Chris S. Haley et Dirk Jan de Koning du Roslin Institute (Ecosse); de Henk Bovenhuis de l'Université de Wageningen (Pays-Bas) à partir des données existantes (héritabilité du caractère) et envisagées (nombre de marqueurs et de familles à produire...). En effet, les simulations ont démontré que la détection de QTL expliquant une faible part de variance (environ 5%) était possible à partir d'un schéma expérimental composé de plusieurs familles F2, d'une stratégie de génotypage sélectif (15 à 20% de la distribution) et d'un caractère présentant une forte héritabilité ($h^2 > 0,4$). Ces résultats ont orienté nos choix de schéma expérimental de production (nombre de familles F2 à produire, etc...) et de phénotypage des individus de la F2.

5.6.2 Jeu de données disponible pour les analyses QTL

Le jeu de données utilisé pour les analyses QTL était composé (1) des phénotypes de 900 individus (2 groupes de phénotype extrême*150 individus*3 familles F2) relatifs aux caractères « survie » et « résistance à l'infection par le virus OsHV1 » (voir chapitre II) et (2) de la cartographie génétique consensus (voir chapitre IV), bâti à partir des données de génotypes de 80 marqueurs au total dont 29 marqueurs SNP, 13 microsatellites développés *in silico* et 38 micorsatellites précédemment publiés (voir chapitre III).

5.6.3 Présentation du logiciel « QTL express »

Suite à la construction d'une cartographie génétique, la recherche de QTL(s) liés à la résistance à la mortalité estivale et à la charge virale individuelle en Herpès virus, a été entreprise. Pour réaliser ces analyses, notre choix s'est porté sur le logiciel QTL express (Seaton et al., 2002). Ce logiciel comporte plusieurs modules d'analyses spécialisés. Dans notre cas (3 familles F2), trois des six modules d'analyses sont applicables au jeu de données. Le but de QTL Express est de fournir des modules faciles à utiliser pour l'analyse de données provenant de populations. Le site est en évolution constante, mais le but ultime est de fournir une gamme d'outils pour analyser les données provenant de différentes structures de population (backcross, F2, etc...). Le logiciel se concentre sur des outils analytiques simples,

mais robustes et puissants. Par conséquent, les méthodes d'analyse de régression des moindres carrés et de variance composante ont été mises en place (Haley et Knott, 1992).

Des outils pour l'analyse par permutation et bootstrapping estiment les régions de confiance pour la localisation des QTL. À l'heure actuelle, les analyses de régression des données ont été appliquées pour la population des structures suivantes:

- (A) une population de type F2 dérivée d'un croisement entre deux populations
- (B) une population de type backcross provenant d'un croisement entre deux populations
- (C) les populations avec un mélange de backcross (réciproque) et d'individus F2 issus de deux populations
- (D) les assemblages de populations de familles de demi-frère
- (E) des familles nucléaires (sib-couples)

5.6.4 Module F2

Ce module est destiné à l'analyse des données à partir d'une population F2 dérivée d'un croisement entre deux lignées. Il est supposé que tout QTL est fixé chez les deux lignées. Par conséquent, il est possible de combiner des informations des QTL chez les différentes familles. Si les QTL ne sont pas fixés, l'analyse peut être menée tout de même, mais une perte de puissance significative sera constatée à mesure que les fréquences des allèles deviennent moins extrêmes (Alfonso et Haley, 1998). Les données de marqueurs sont fournies pour les grands-parents (F0) et la descendance F2, et les phénotypes sont fournis pour les individus F2. La méthode a été développée pour les lignées par Haley et Knott (1992) et prorogée pour les lignées de types « outbred » par Haley et collaborateurs (1994). L'analyse procède en deux étapes. Les premières informations sur la position des marqueurs, ainsi que les génotypes des marqueurs sont utilisés pour calculer des probabilités d'individus héritant 0, 1 ou 2 allèles de chacun des deux fondateurs des lignées à travers les positions du génome et le parent d'origine probable des allèles. Dans la deuxième étape de l'analyse des données phénotypiques sont régressées sur ces coefficients. La simple nature de l'approche de régression signifie qu'il est possible d'utiliser différents modèles génétiques et environnementaux, tels que la présence d'un ou de deux QTL liés, des effets additifs, de dominance et du parent d'origine, l'effet des QTL, les effets des facteurs environnementaux (effets fixes) ou covariables, les interactions avec des QTL (effets fixes), etc.

La ségrégation de distorsion est calculée comme une déviation par rapport à la moyenne observée des coefficients des effets additifs, de dominance et du parent d'origine sous l'hypothèse nulle de la ségrégation mendélienne. Un test statistique est effectué en calculant la différence entre moyenne observée et attendue et une estimation de l'erreur standard de cette différence. Sous l'hypothèse nulle d'absence de distorsion de ségrégation, ce ratio suit comme une distribution t . Le graphique avec la ségrégation distorsion contient deux lignes horizontales à $t = -2,0$ et $t = +2$, correspondant à la valeur nominale P-valeurs d'environ 0,05. Le niveau d'information est calculé suivant le modèle de Knott et collaborateurs (1998).

Le test par permutation (Churchill et Doerge, 1994) est mis en oeuvre par la permutation des coefficients de la matrice choisie pour un chromosome donné chez un individu. De cette manière, les estimations des effets fixes ne sont pas affectées, alors que la relation entre le génotype et le phénotype sont permutés. Pour estimer l'intervalle de confiance d'un emplacement QTL, un ré-échantillonnage de type bootstrap est disponible en option.

5.6.5 Module “Half-Sib”

Cet outil d'analyse est consacré à l'analyse des familles non consanguines et étudie la ségrégation des allèles de chacun des deux parents (F1) chez la descendance (F2). Les génotypes des descendants servent à déterminer les haplotypes parentaux. Le parent commun est censé apporter des informations génotypiques uniquement. Ces haplotypes doivent être connus et ne comporter aucune erreur. Pour chaque descendant, la probabilité d'hériter d'un des haplotypes est calculée dans chaque intervalle en utilisant l'information provenant des marqueurs flanquants composant ce même intervalle. Suite au travail de Neimann-Sorensen et Robertson (1961) sur l'analyse du groupe sanguin par des marqueurs simples, la méthode a été développée par Knott et Haley (1992), et a été largement utilisée pour l'analyse des groupes de demi-frères chez les bovins laitiers et les porcs (de Koning et al., 1998). L'analyse suit une procédure en deux étapes identique au module F2.

Dans ce modèle, les effets des QTL ne sont pas supposés être identiques chez chacune des familles. Cependant, si un QTL existe, celui-ci doit ségréger chez plus d'une famille (Knott, 2005). Le test du seuil de significativité du QTL est basé sur un test F (F ratio) calculé de manière identique au test appliqué dans le module F2.

Enfin, l'intervalle de confiance est calculé selon la méthode de bootstrapping au sein de chacune des familles de telle sorte que le nombre de descendants par parent en commun reste constant.

5.6.6 Module d'analyse en intra-famille – module « Tree »

Le module « Large Single Full-Sib Family Analysis ou Tree » a été utilisé parce qu'il est conçu pour le cas d'une famille avec deux parents qui ne sont pas supposés provenir d'un croisement entre deux lignées génétiques distinctes et différentes. Avec ce module, trois estimations peuvent être obtenues : l'estimation de la composante paternelle, maternelle et de l'interaction des deux premières composantes. En effet, l'analyse établit une comparaison entre les deux gamètes mâles acheminés par le parent mâle (composante mâle), et les deux gamètes transportés par le parent femelle (composante maternelle).

L'information provenant des grands-parents est utilisée pour aider à identifier les deux gamètes parentaux transmis. Si la composante paternelle est importante, cela suggère que les deux gamètes paternels portaient des allèles d'effets différents pour les QTL. Pour estimer ces trois composantes, quatre modèles différents sont disponibles sous QTL express: pat (pour estimer la composante paternelle seulement), mat (estime la composante maternelle seulement), pat + mat (estimation à la fois les composants paternel et maternel) et, enfin, pat + mat + int (pour estimer la composante interaction).

Afin de rétablir la distribution normale des données de phénotype, et plus particulièrement pour le caractère « charge virale », les données individuelles de charge, exprimées en nombre de copies d'ADN viral par mg de tissu frais, ont été transformées en Log_{10} . Les données liées au phénotype « mortalité » sont des données binaires (0 ou 1) et ont été conservées comme tel.

La recherche de QTL a été entreprise de manière identique mais séparée pour chacun des dix groupes de liaison identifiés au cours de la construction de la cartographie génétique. Pour chacun des modules, les paramètres d'analyses sont les suivants :

- Module F2 : Modèle (a+d) ; Analyse : 1 QTL ; step = 1cM ; Bootstrapping avec rééchantillonnage ; 1000 itérations ; 2 caractères
- Module Half-sib : Dam/Sire as common parent; Modèle (a+d) ; Analyse : 1 QTL ; step = 1cM ; Bootstrapping avec rééchantillonnage ; 1000 itérations ; 2 caractères
- Module Tree : Modèle (pat+mat) ; Analyse : 1 QTL ; step = 1cM ; Bootstrapping avec ré-échantillonnage ; 1000 itérations ; 2 caractères
- Seuil de signification placé à un Lod score égal à 3 pour chaque chromosome

5.7 Résultats

5.7.1 Module « F2 »

La première analyse de chacun des dix groupes de liaisons a révélé la présence significative de cinq régions QTLs situées dans les groupes de liaisons II, V, VI, VII et IX (Annexe 1 et figure 35). Suite à ces premières analyses, la recherche de QTL a été répétée pour le groupe de liaison VI où deux pics QTL, apparemment significatifs, apparaissent mais en utilisant le modèle de détection à deux QTL et en fixant l'effet du QTL, situé à 63cM, en tant qu'effet fixe (genetic background). Cette modification a révélé la présence d'un second QTL au sein de ce même groupe de liaison situé à 112 cM (figure 36). Puis, la recherche de QTL a été répétée de la façon suivante : tour à tour, les effets de chacun des QTL ont été fixés en tant que cofacteur de l'analyse et la recherche de QTL a été répétée. Puis, pour chaque groupe de liaison, la recherche de QTL a été effectuée en fixant l'effet d'un, puis de deux, puis des trois groupes de liaisons simultanément. Cette analyse a mis en évidence un faux positif. En effet, lorsque les effets des quatre régions QTL, des groupes de liaison V, VI, VII et IX, sont fixés en tant que cofacteur, l'analyse du LG II seul ne met plus en évidence le QTL précédemment identifié dans ce groupe de liaison (figure 37).

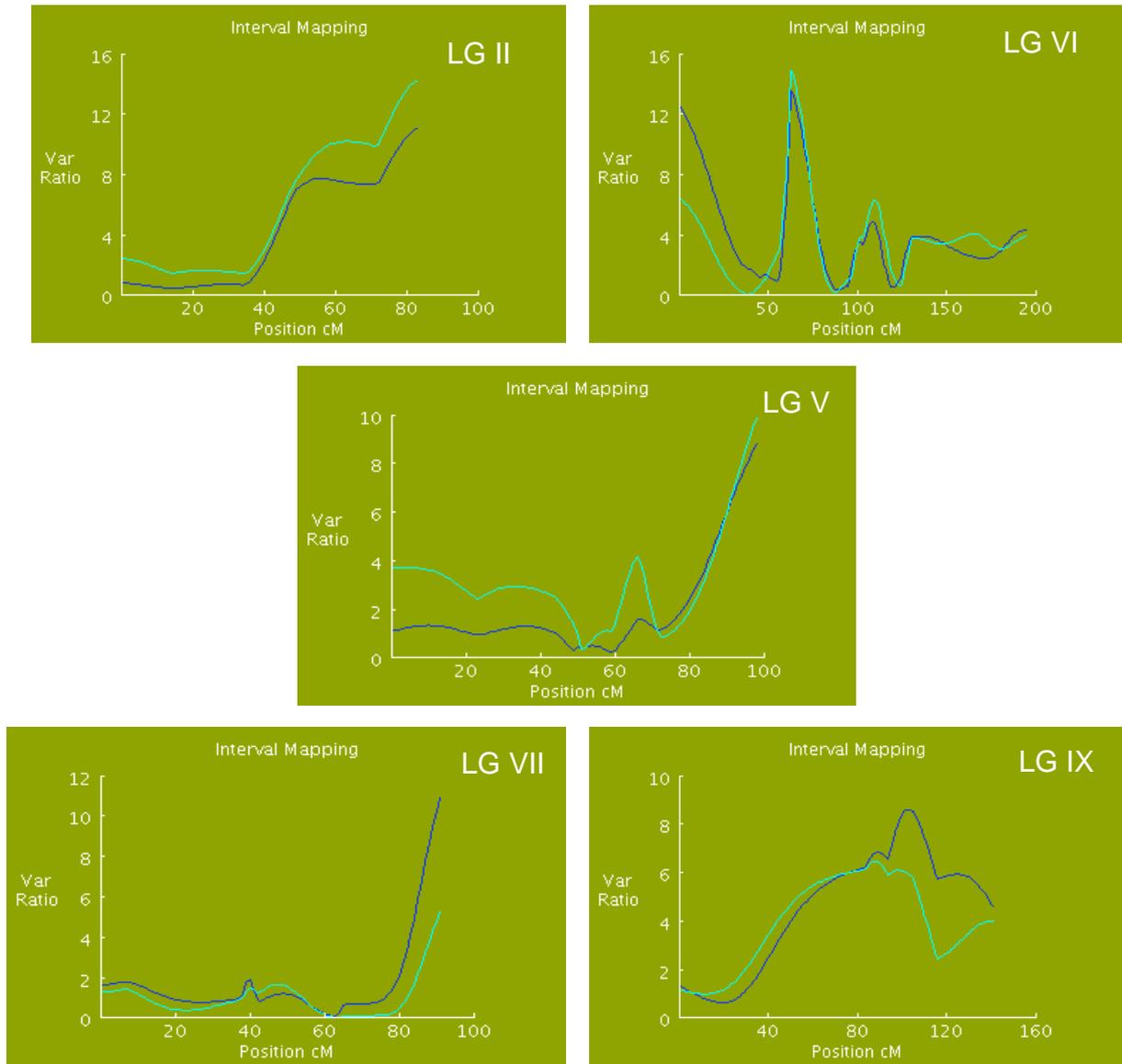


Figure 35 : Primo-détection de QTL dans les groupes de liaison II, V, VI, VII et IX

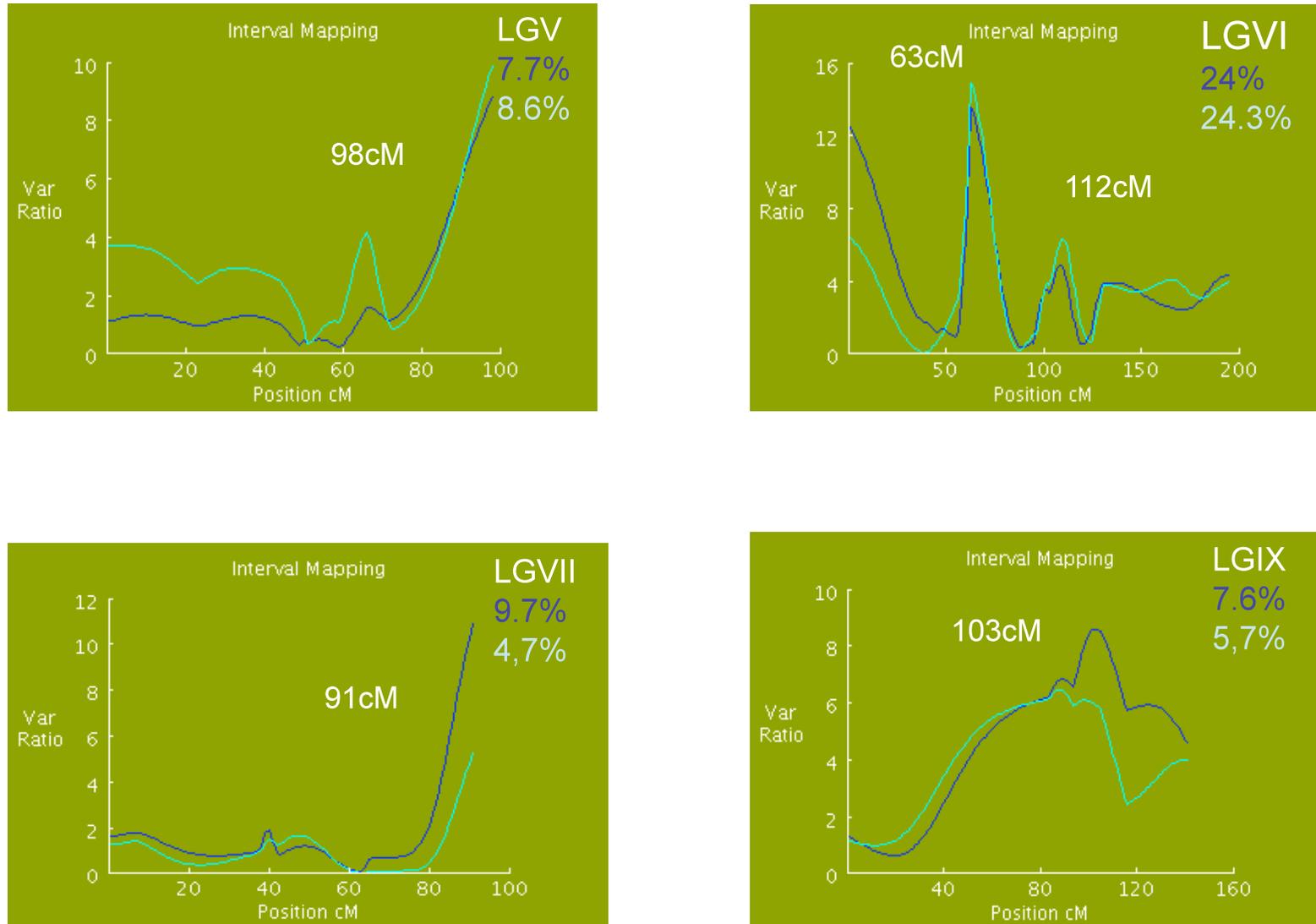


Figure 36 : Localisation des cinq QTL dans les groupes de liaison V, VI, VII et IX

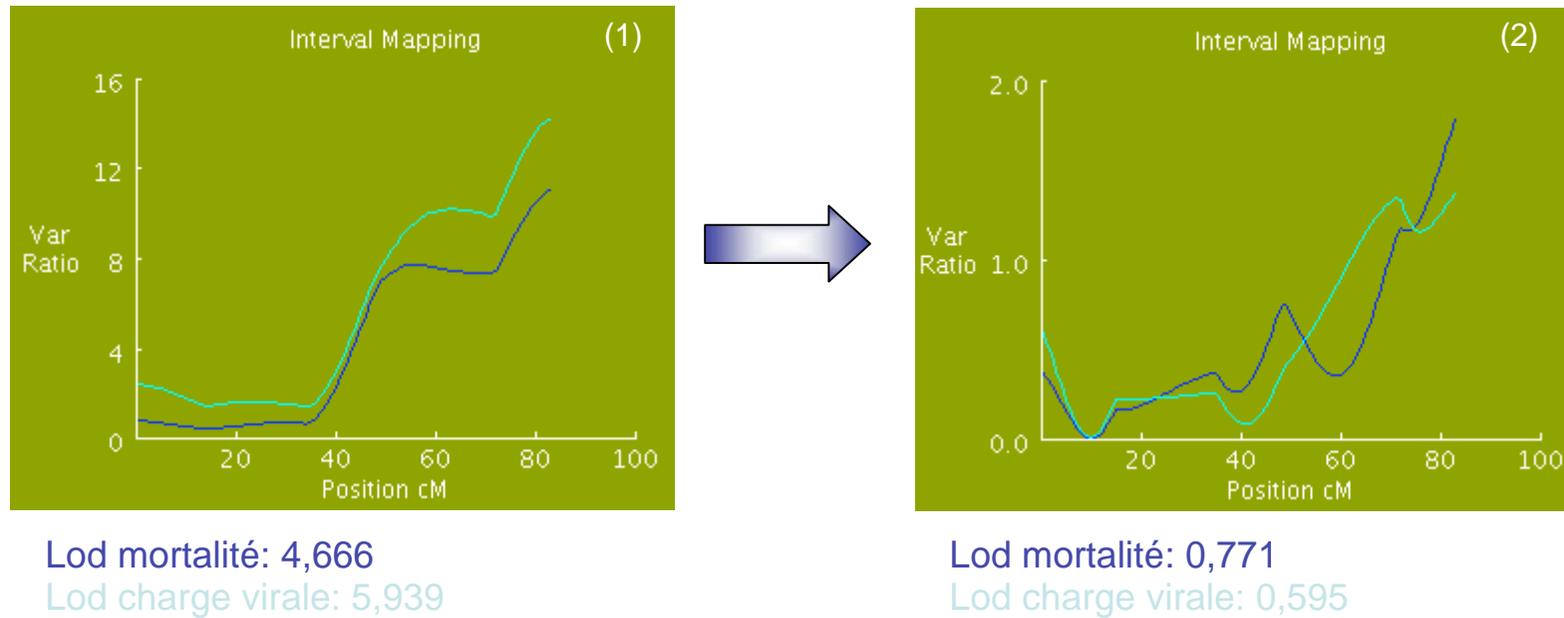


Figure 37 : Mise en évidence d'un faux QTL au sein du groupe de liaison II

(1) Aucun effet de QTL fixé

(2) Effets des QTL des LG V, VI, VII et IX fixés

En utilisant le module « F2 », la recherche de QTL met en évidence cinq régions QTLs positionnées dans les groupes de liaison V, VI, VII et IX. La région QTL située sur le groupe de liaison II, a été éliminée. Les positions relatives de chacune de ces régions ainsi que leur intervalles de confiance à 95% et la part respective de variance pour les deux caractères phénotypiques sont résumés dans les tableaux 18 et 19.

Tableau 18 : Part de variance phénotypique respective expliquant la variation des deux caractères.

Groupe de liaison	Mortalité		% de variance	Charge Virale		% de variance
	Modèle Complet	Modèle Réduit		Modèle Complet	Modèle Réduit	
LGV	220,67	225	7,7	5702,87	5828,61	8,6
LGVI	211,48	225	24,0	5474,41	5828,61	24,3
LGVII	214,64	219,99	9,7	5621,12	5688,59	4,7
LGIX	219,51	223,74	7,6	5686,96	5769,53	5,7
Total			49,0			43,4

Tableau 19 : Lod score, positions estimées et intervalles de confiance à 95% des régions QTL pour chaque caractère.

Groupe de liaison	Mortalité			Charge Virale		
	Position (cM)	Lod score	IC à 95%	Position (cM)	Lod score	IC à 95%
LGV	91,9	7,621	21- 98	80	6,411	0 - 98
LGVI - QTL1	63	12,061	0 - 110	66,7	12,198	0 - 135
LGVI - QTL2	112					
LGVII	89	6,031	51 - 91	72	3,137	3 - 91
LGIX	103	3,744	42 - 140	93,7	3,176	15 - 141

Les pourcentages totaux de variance expliquée par les cinq régions QTL identifiées sont estimés à 49 et 43,4% respectivement pour le caractère de survie à la mortalité estivale et le caractère charge virale individuelle en Herpès Virus. Pour le premier caractère, le pourcentage le plus important, est de 24% pour les deux QTL identifiés dans le groupe de liaison VI, alors que le plus faible est de 7.6%, pour le QTL identifié dans le groupe de liaison IX. De même, pour le caractère charge virale individuelle en Herpès virus, la plus grande part de variance phénotypique (24.3%) est attribuée aux deux régions QTL du groupe de liaison VI tandis que la plus faible part de variance (5.7%) est attribuée au QTL du groupe de liaison IX. De plus, l'intervalle de confiance à 95% de la position des QTL est relativement large avec un intervalle maximum de 110cM (LG VI) et un intervalle minimum de 40cM (LGVII).

Les effets de dominance et additif des QTL sont forts pour le caractère « survie » pour le QTL du LGV. Pour le LG VI, les effets des deux régions QTL sont de sens opposés pour les deux caractères alors que dans le LG IX, l'effet de dominance est de sens opposé à l'effet additif pour les deux caractères. L'ensemble des résultats est regroupé en annexe 2.

Le phénotype « charge virale individuelle » a été fixé en tant que covariable de l'analyse. Cela signifie que l'effet de ce phénotype est soustrait à l'effet du second phénotype. Cette analyse révèle que l'ensemble des parts de variance des cinq régions QTL n'est pas uniquement lié au phénotype « charge virale individuelle » car une part de variance uniquement lié au phénotype « survie » est mise en évidence (figure 38).

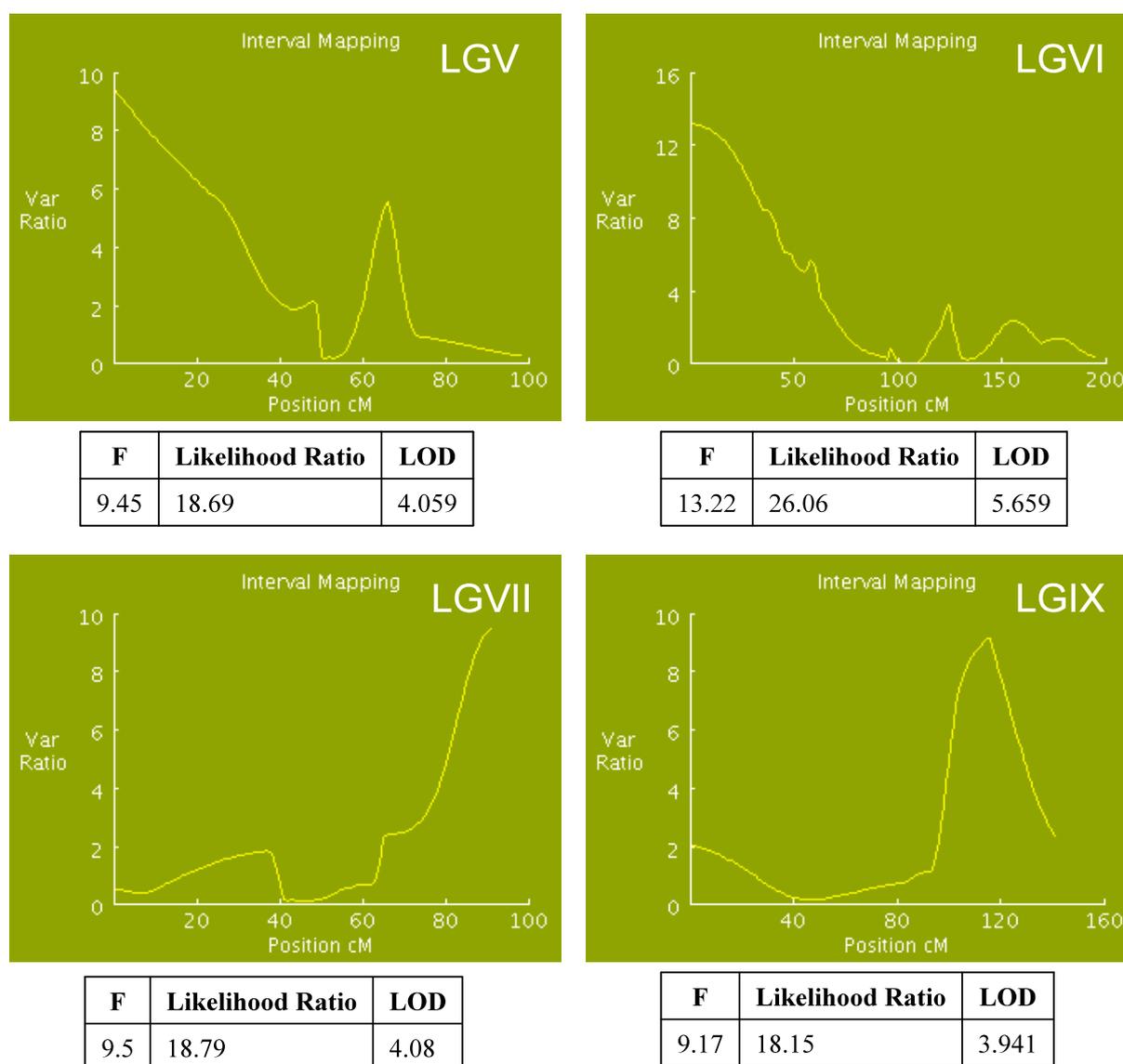


Figure 38 : Part de variance phénotypique lié au phénotype « survie ».

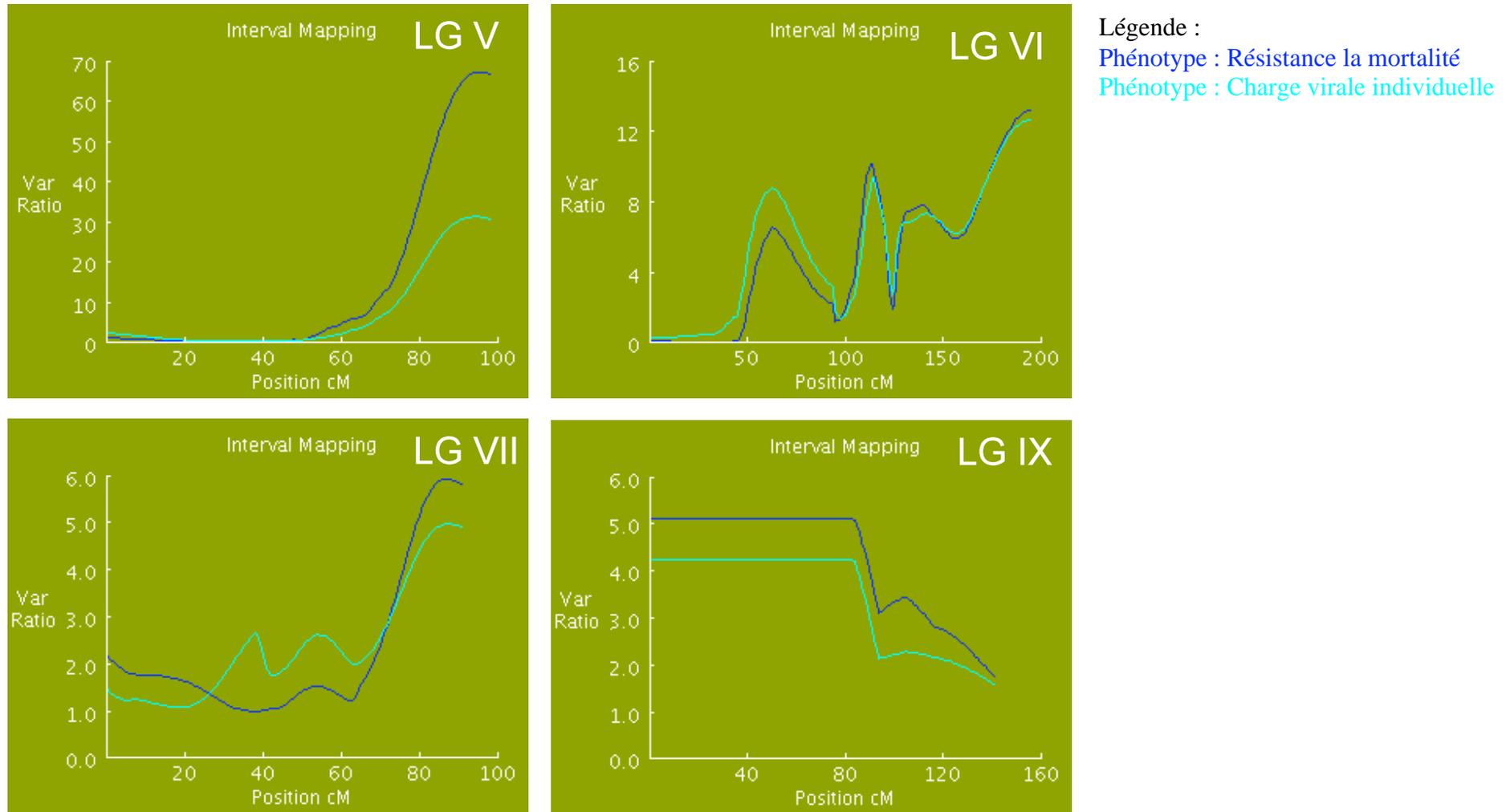


Figure 39: Composante phénotypique maternelle (Dam effect - Module Half-Sib)

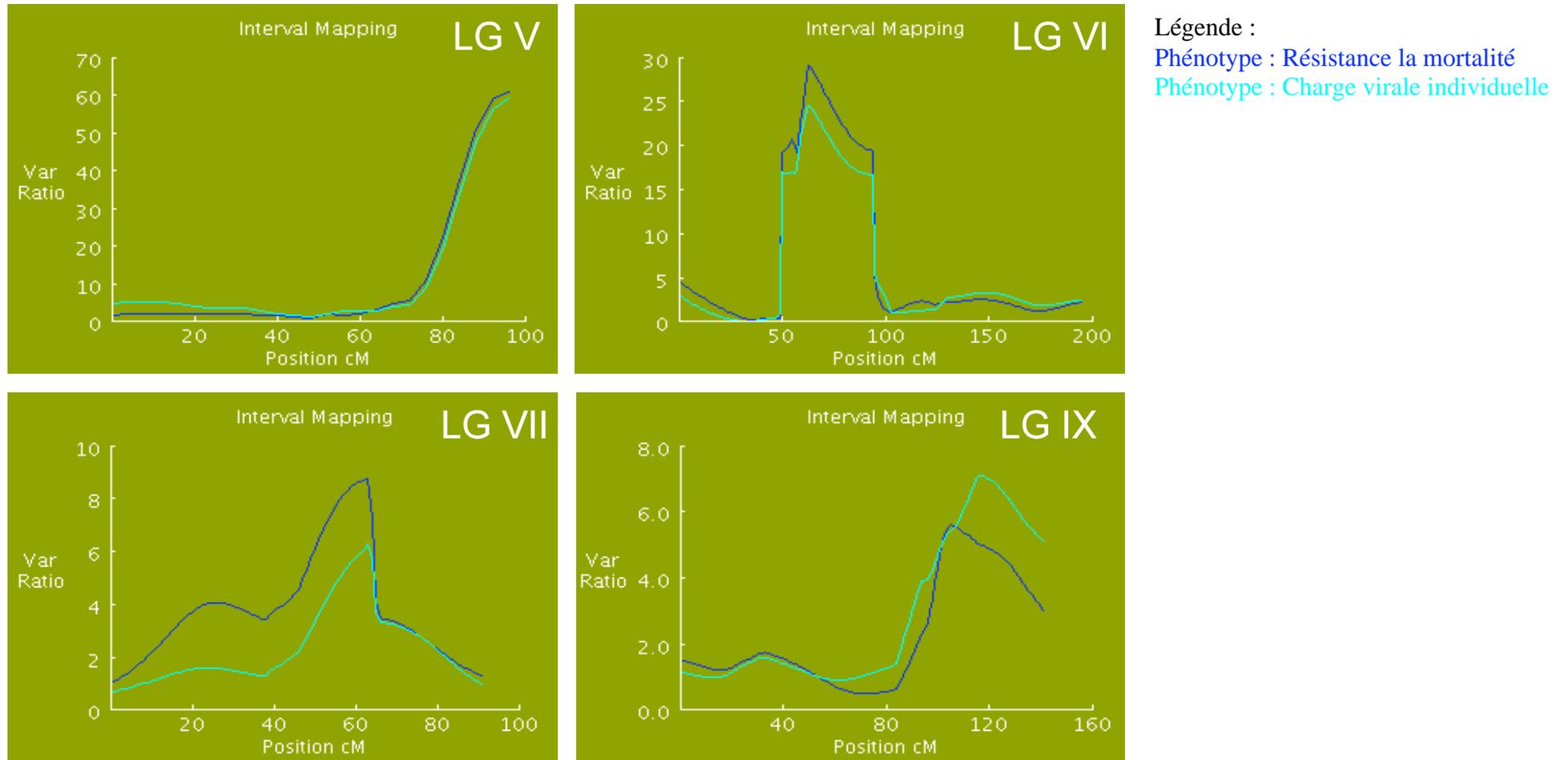


Figure 40: Composante phénotypique paternel (Sire effect - Module Half-Sib)

5.7.2 Module « Half sib »

Pour rappel, la méthode d'analyse de ce module Half-Sib, estime l'effet des QTL chez les individus F1 par sexe et indépendamment de l'origine des allèles. Les résultats résumés en annexe 3, et représentés en figures 39 et 40, montrent des effets paternels et maternels très forts avec une probabilité (Likelihood ratio) maximale de 181,92 (Groupe de liaison V, composante maternelle) à une probabilité minimale de 12,69 (groupe de liaison IX, composante maternelle). Les composantes paternelles sont significativement supérieures aux composantes maternelles pour les deux caractères phénotypiques (test de Wilcoxon, 2 distributions appariées, $Z_{mortalité} = 0.465$ et $Z_{Charge\ Virale} = 0,068$).

5.7.3 Module « Large single full-Sib family analysis- Tree »

Ce module estime le contraste, au sein d'une unique famille, entre les marqueurs hérités du parent mâle ou du parent femelle. Les résultats montrent une ségrégation différentielle entre les trois familles F2 (E*T, R*O et Z*AC) des 5 régions QTL précédemment mises en évidence par le module de recherche F2 (Annexe 4 et figure 41). En effet, les seuls QTL significatifs (Likelihood Ratio > F ratio et Lod score >3) chez la famille E*T sont les QTL des groupes de liaison VI et VII ; chez la famille R*O, les QTL des groupes de liaison V, VI et VII ; et chez la famille Z*AC le QTL du groupe de liaison IX.

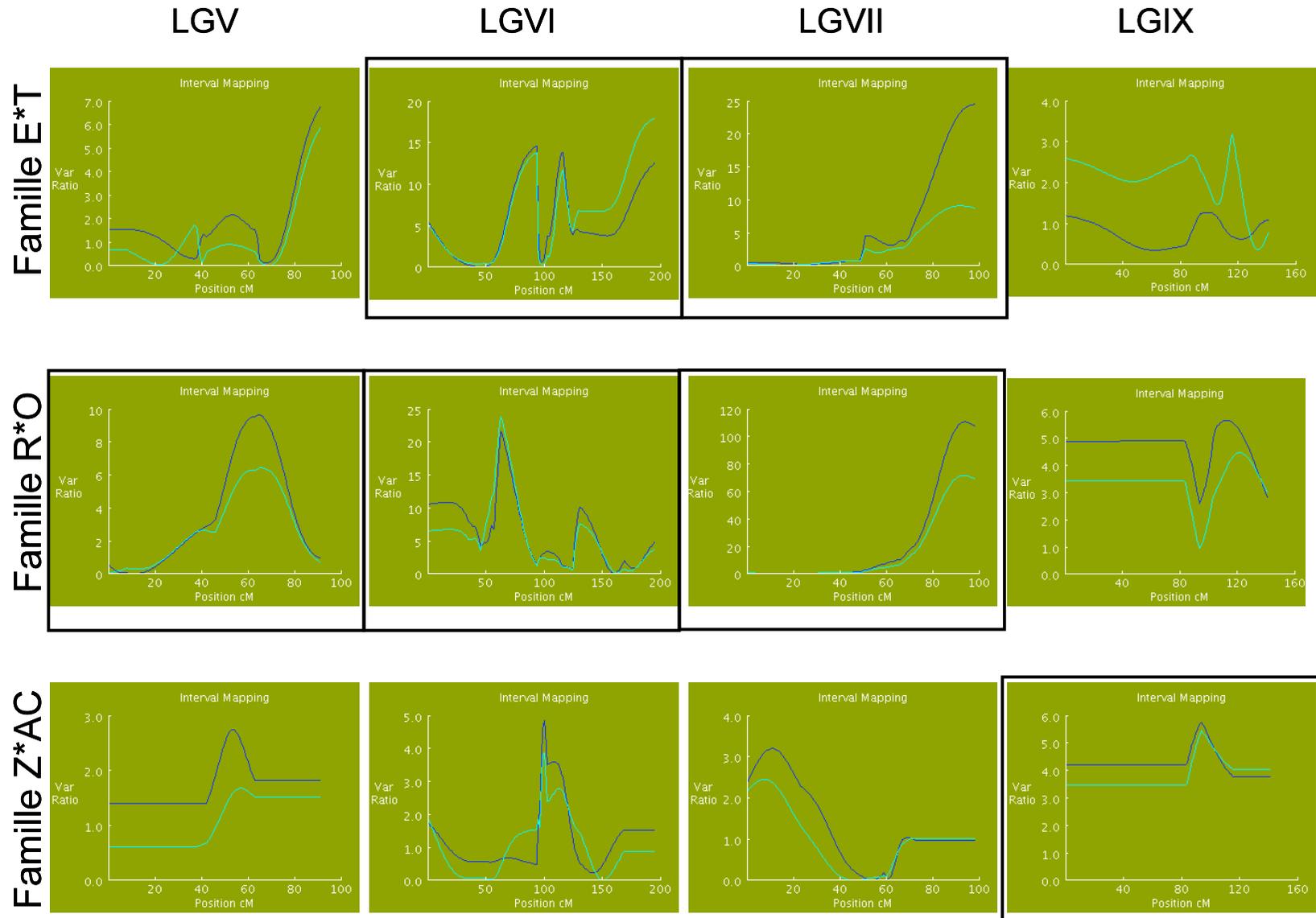


Figure 41 : Ségrégation différentielle des QTL dans les familles F2 (les QTL significatifs sont encadrés)

5.8 Discussion

5.8.1. Nombre de QTL identifié et part de variance expliquée

Le modèle de détection de QTL de type « F2 » nous a permis de mettre en évidence cinq QTL. Ces régions chromosomiques expliquent une large part de variance des deux caractères étudiés. L'analyse du groupe de liaison VI a mis en évidence deux QTL situés respectivement à 63 et 112 cM. Ceux-ci sont distingués l'un de l'autre pour plusieurs raisons : (1) la distance entre ces deux QTL est suffisamment importante pour ne pas confondre ces deux régions QTL avec un unique QTL relativement « large », c'est-à-dire qui s'étend sur une longue distance génétique ; (2) la fixation de l'effet d'un des deux QTL, met en évidence le second et inversement ; (3) l'analyse individuelle des familles F2, au travers du module « tree », révèle également la présence de deux pics QTL significatifs, pour les deux caractères, chez les familles R*O et E*T à des positions relativement similaires à celle obtenues dans le module F2 (1^{er} pic au tour de 60 cM et 2nd pic aux alentours de 120 cM ; figure 33). Ce résultat confirme donc la présence de deux régions chromosomiques distinctes liées à la variance phénotypique des deux caractères dans le groupe de liaison VI.

Cependant, l'estimation de l'intervalle de confiance de ces deux régions chromosomiques est le problème principal. En effet, cet intervalle est relativement large (110 cM) et s'explique par le recouvrement des intervalles respectifs de chacun des deux QTL. La cartographie d'un plus grand nombre de marqueurs moléculaires dans ce groupe de liaison permettra de réduire et d'estimer plus justement les intervalles de confiance de chacun des deux QTL détectés. Sur un croisement entre deux lignées fixées, Darvasi (1997) montre que la sélection pour le génotypage de marqueurs supplémentaires de 40 à 50% des individus, présentant des performances extrêmes pour le caractère, concerné par le QTL, permet d'atteindre la taille minimum de l'intervalle de confiance.

La fixation successive puis simultanée des effets des QTL significatifs a révélé un QTL, dans le groupe de liaison II, que nous avons considéré comme faux positif. L'hypothèse la plus probable pour expliquer une telle observation est la suivante : la stratégie de génotypage sélectif tend parfois à créer des corrélations ou liaisons entre marqueurs, même si ceux-ci sont indépendants. Ainsi, la part d'individus très sensible à la mortalité, et comparons la à la part d'individus les plus résistants à ce même phénomène. Dans ce cas, les individus les plus sensibles vont porter la majorité des allèles caractéristiques de la sensibilité, alors qu'à

l'inverse, les individus les plus résistants vont porter la majorité des allèles caractéristiques de la résistance. Cela génère alors une fausse liaison que l'on peut interpréter comme une région QTL significative (Haley, C ; comm. pers.).

Les parts de variance expliquée sont relativement importantes avec, au total, près de cinquante pourcents de variance expliquée pour le caractère phénotypique « résistance à la mortalité estivale ». Ces pourcentages peuvent être surestimés pour deux raisons : (1) la stratégie de génotypage sélectif ; (2) la sélection des QTL significatifs ce qui signifie que les QTL reflètent également la part de variance liée à des régions QTL à effets plus faible qui n'ont pas été détectées . En effet, comme décrit précédemment, le génotypage a été effectué sur 30% de la distribution chez chacune des trois familles F2. Les modèles de recherche de QTL tendent à surévaluer cette part de variance (Lander et Botstein, 1989; Knott, 2005). Les travaux de simulation réalisés par Ronin et collaborateurs (1998) montrent que le biais, introduit par la stratégie de génotypage sélectif, est proportionnel à la part de variance expliquée par le QTL. Pour un QTL dont la part de variance est supérieure à 10%, les auteurs recommandent de génotyper 50% de la distribution. Darvasi et Soller (1992) ont estimé le biais introduit par la stratégie de génotypage pour l'effet additif par la formule simplifiée : $(1+xi)$ où x est l'ordonnée de la valeur moyenne du caractère pour une distribution normale (qui dépend de p , la proportion de la distribution génotypée) et i , la valeur moyenne attendue pour le caractère chez les individus composant le phénotype extrême supérieur. Pour des caractères corrélés, le biais peut être estimé selon la formule $raxi$ (Bovenhuis et al., 2000), sous condition que les corrélations génotypiques et phénotypiques soient identiques ; où $r = (1-2\theta)$, avec θ le taux de recombinaison entre le marqueur et le QTL; a l'effet additif et $xi = [1+Zi]$ avec Z un estimateur de la déviation par rapport à la valeur moyenne du caractère pour une distribution normale et i un estimateur de la sélection égal à $p/2$ avec p , la proportion de la distribution génotypée. Cependant, les modèles de cartographie de QTL qui intègrent ce biais doivent être affinés (Lin et al., 1996). Ensuite, ces mêmes modèles ne détectent pas les QTL à effets faibles pour plusieurs raisons (ex : faible densité en marqueurs moléculaires de la carte de liaison, taille de l'échantillonnage, stratégie de génotypage sélectif). Il existe donc une part de variance au niveau de ces QTL à effets faibles qui n'est pas mise en évidence par l'analyse effectuée par le module F2. De plus, la ségrégation de distorsion, mise en évidence au cours du processus de construction des cartes de liaison, ainsi que la pression de sélection, subie au cours du programme d'obtention des animaux (Hackett et Broadfoot, 2003; Lynch et Walsh, 1998), introduisent un biais qui tend également à surévaluer les pourcentages de part de

variance attribués à chacune des régions QTL. Néanmoins, les résultats obtenus permettent une primo-localisation de 5 régions QTL et l'estimation de leurs effets.

5.8.2. Ségrégation différentielle entre les familles F2

Ce type d'analyse a mis en évidence une ségrégation différentielle des cinq régions QTL parmi les trois familles F2. Cette ségrégation différentielle peut avoir deux type de causes : (1) tous les marqueurs moléculaires ne soit pas également informatifs entre les familles et (2) tous les QTL identifiés ne ségrégent pas dans l'ensemble des familles. De plus, les phases de liaison QTL / marqueurs flanquants, associant différents allèles au locus QTL aux allèles des marqueurs flanquants entre les familles. Cela signifie que toute association entre un marqueur et un QTL se fait au niveau de la famille plutôt qu'au niveau de l'ensemble des familles. En effet, le déséquilibre de liaison, entre le marqueur et le QTL lié, est généré au sein des familles, mais il peut ne pas exister dans toutes les familles. De plus, d'autres différences entre familles qui sont causés par d'autres loci affectant le caractère d'intérêt existent (Knott, 2005). Ces résultats soulignent l'importance de multiplier le nombre de familles F2 à produire pour augmenter la probabilité d'observer la ségrégation d'un QTL mais aussi la complexité des résultats qui en découlent. Cette probabilité est d'autant plus forte que l'héritabilité du caractère est forte, mais elle est réduite si les QTL ne sont pas fixés dans des lignées parentales, ce qui est le cas dans notre schéma expérimental.

Les premiers travaux d'identification d'association entre marqueurs et génotypes, chez les ostréidés, ont été menés chez l'huître creuse américaine, *Crassostrea virginica* par Yu et Guo (2006). Le phénotype considéré était la résistance à deux pathogènes (MSX et *Perkinsus marinus*), qui affectent les rendements de production. Les analyses menées ne sont pas de véritables analyses QTL, au sens premier du terme, car celles-ci sont basées sur l'analyse du changement de fréquence alléliques entre un échantillon d'individus prélevé avant l'événement de mortalité, et un second prélevé après cet événement. Néanmoins, 12 régions associées à la résistance à ces deux pathogènes ont été mises en évidence. Yu et Guo (2006) affirment que la principale difficulté, dans la recherche de QTL liés à la résistance à un pathogène, réside dans la quantification du phénotype. Nos travaux ainsi que des travaux de détection de QTL liés à la résistance à la bonamiose (organisme pathogène) chez l'huître plate européenne, *Ostrea edulis* (Lallias, 2006), modèrent cette affirmation. En effet, de nouveaux outils, tels que la quantification de l'ADN d'un virus par PCR et un schéma expérimental

adapté (suivi de mortalité quotidien en raceway) permet de récupérer un fragment de tissu en cours de dégradation et ne limite donc pas les analyses QTL.

5.8.3. Relations entre QTL de survie et de charge virale

Il est intéressant de noter que les QTL, mis en évidence pour les deux caractères phénotypiques étudiés sont co-localisés. Les positions et les intervalles de confiance sont très similaires, voire identiques, entre les deux caractères pour chaque QTL. Dans le cadre de notre expérience de suivi de mortalité, nous pouvons supposer que l'architecture génétique des deux caractères est très similaire. L'étroite corrélation phénotypique précédemment mise en évidence, entre l'intensité de la mortalité (ou nombre d'individus morts par famille F2) au cours du suivi estival et la charge virale individuelle (voir chapitre II), tend à appuyer cette hypothèse. Néanmoins, l'analyse du phénotype « survie », en plaçant le caractère « charge virale individuelle » en tant que cofacteur, montre que toute la variation phénotypique n'est pas uniquement expliquée par ce dernier caractère. Nous pouvons en supposer que (1) l'infection virale n'est pas la seule cause de mortalité des individus au cours de ce suivi de mortalité estivale car la part de variance expliquée par ce seul caractère est compris entre 4 et 10% selon le QTL ou (2) qu'il existe une différence de puissance d'analyse entre les deux caractères puisque l'un est binaire et l'autre quantitatif.

Cette observation dans une certaine mesure est en accord avec les résultats obtenus au cours du programme MOREST, au cours duquel, la pluralité des causes de mortalité chez l'huître creuse, durant la saison estivale, a été mise en évidence (Samain et al., 2007). Néanmoins, au cours de ce programme, la recherche du virus OsHV1 n'avait pas été systématisée, ce qui ne nous permet pas d'établir un parallèle clair entre les causes de mortalité des lignées sélectionnées ayant servi de base à cette étude et les causes de mortalité des individus F2 de notre expérience. De plus, suite à notre expérience, il apparaît difficile d'affirmer que l'infection par l'herpès virus est la seule cause de mortalité, la recherche d'autres agents pathogènes, notamment de type *Vibrio*, ou d'autres facteurs n'a pas été menée. Le lien entre survie et charge virale dans notre expérience reste néanmoins très fort. Une telle corrélation entre la mortalité et la présence d'un agent pathogène a également été démontrée, au sein du genre *Crassostrea*, notamment avec MSX et *Perkinsus marinus* chez *C. virginica* (Yu et Guo, 2006).

5.9 Perspectives

Les perspectives à l'issue de cette première identification de QTL liés à la mortalité estivale et à la charge virale OSHV1 sont nombreuses. L'intervalle de confiance de la position des QTL s'étend sur plusieurs dizaines de centimorgans, incluant des centaines de gènes. La cartographie du déséquilibre de liaison permettra de réduire cet intervalle. Néanmoins, cet intervalle réduit contiendra toujours plusieurs dizaines ou centaines de gènes. Cela rend encore difficile l'intégration des QTL dans des programmes de sélection assistée par marqueurs. Afin d'atteindre cet objectif, il faut tout d'abord réduire les zones QTL à un intervalle de 5 à 10 cM en intégrant un plus grand nombre de marqueurs moléculaires sur les cartes de liaison, puis identifier du polymorphisme responsable de la variation du caractère (ou QTN). L'identification et la vérification de QTN a été menée chez trois espèces : gène *DGATI* chez le bétail laitier (Grisart et al., 2002) ; gène *IGF2* chez le porc (Van Laere et al., 2003) et le gène *GDF8* chez le mouton (Clop et al., 2006). Ce processus est relativement long et difficile car ces travaux ont nécessité plus de 4 ans de travaux (Ron et al., 2007). L'ensemble des grandes étapes de ces travaux est schématisé en figure 42.

Le développement des ressources génomiques chez *C. gigas* va accélérer ce type de processus de cartographie fine des QTL et de validation de QTN (Figure 42). Chez l'huître creuse, certaines « briques » viennent d'être récemment posées. En effet, de nombreux éléments tels que des marqueurs moléculaires (Li et al., 2003; Sekino et al., 2003; Yamtich et al., 2005; Sauvage et al., 2007; Yu et al., 2007; Yu et al., 2008); des cartographies génétiques (Hubert et Hedgecock, 2004; Li et Guo, 2004), des gènes candidats identifiés par des travaux d'expression de gènes (Huvet et al., 2004; Tanguy, 2004; Jenny et al. 2007; Fleury et al., 2008) sont désormais disponibles. A moyen terme, les travaux doivent se concentrer sur le clonage positionnel et la cartographie physique. Enfin, à long terme, un schéma d'introgession de QTL, assisté par marqueurs, doit être bâti pour envisager des essais fonctionnels et boucler le schéma général de sélection assistée par marqueurs.

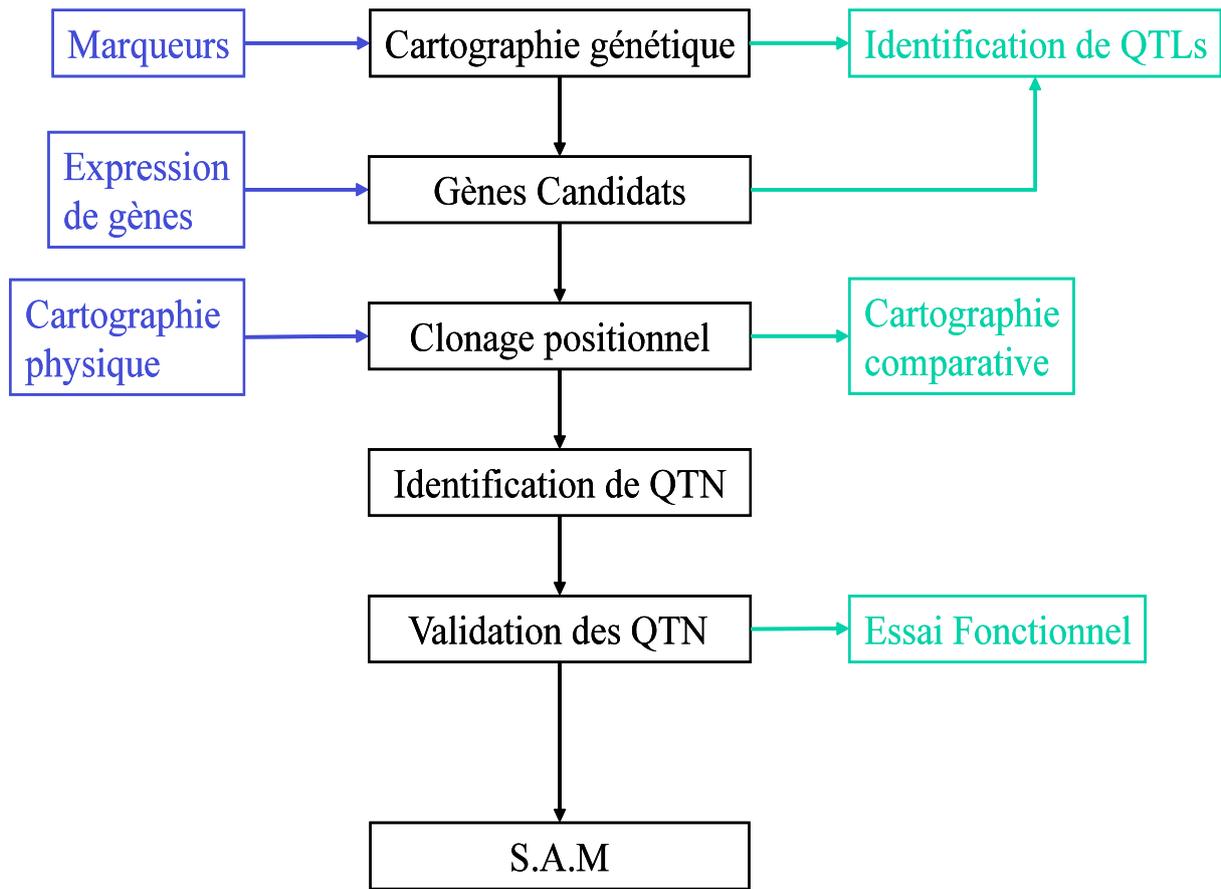


Figure 42 : Schéma conventionnel de validation des QTN depuis les QTL (Ron et al., 2007).

CHAPITRE VI

CONCLUSION GENERALE

ET PERSPECTIVES

Les présents travaux ont permis d'approfondir les connaissances sur le génome de l'huître par le développement et la caractérisation de marqueurs moléculaires de type SNP, de bâtir une carte génétique consensus qui a servit de base à la localisation de régions chromosomiques associées à la variance des caractères de survie et de résistance à l'infection par le virus OsHV1, qui expliquent une importante part de variance de ces deux caractères et enfin de décrire la dynamique d'un évènement de mortalité associée au virus OsHV1.

De plus, un ensemble de marqueurs microsatellites identifiés dans des séquences EST ont été développés. Ce type de marqueur est très intéressant et utile pour un grand nombre d'applications, notamment chez *C. gigas*. Tout d'abord, ces marqueurs sont des marqueurs très polymorphes et donc très informatifs dans la construction de cartes de liaison. C'est pourquoi, un grand nombre de cartographies sont basées sur ce type de marqueurs, notamment chez les poissons (Chistiakov et al., 2006). Ensuite, ce type de marqueurs peut être utilisé dans des analyses génétiques de population naturelles ou pour la gestion de stocks (Schlotterer et Goldstein, 1999) mais aussi pour l'identification d'individus et l'assignement de parenté entre individus (Taris, 2005). D'un point de vue plus fondamental, ces marqueurs sont couramment utilisés en phylogénie et conservation de la diversité génétique (Goldstein et al., 1999). De plus, les marqueurs moléculaires de type microsatellite sont très bien adaptés à la recherche de QTL de part leur large répartition au sein des génomes et leur haut niveau de polymorphisme (Glazier et al., 2002). Le développement et l'utilisation de marqueurs microsatellites chez *C. gigas* sont des perspectives intéressantes qu'il faut poursuivre car ces marqueurs seront utilisés dans les différents domaines précédemment décrits.

6.1 Poursuite des travaux de cartographies

A notre connaissance, les marqueurs SNP développés au cours de cette thèse sont les premiers de ce type publiés et cartographiés chez l'huître creuse, *Crassostrea gigas*. La détection du polymorphisme nucléotidique est relativement aisée car le niveau de polymorphisme nucléotidique, qui est l'un des plus élevés décrits dans la littérature, facilite cette étape du développement de ces marqueurs. L'augmentation récente du nombre de séquences EST disponibles chez *C. gigas* (Tanguy et al., 2008), devrait également permettre d'identifier de nombreux nouveaux sites polymorphes, notamment pour des gènes impliqués dans les mécanismes de résistance à la mortalité estivale mis en évidence par des analyses en microarrays (Jenny et al., 2007; Fleury et al., 2008). Ces gènes pourront être intégrés, *a posteriori*, dans une carte génétique plus dense.

Cependant, cette relative facilitée de détection des SNP est contre balancée par les difficultés techniques que nous avons rencontrés pour réaliser le génotypage en « haut-débit » de ces marqueurs, en collaboration avec la plate-forme du Max Planck Institute de Berlin. La richesse en bases GC et le niveau de polymorphisme nucléotidique élevé deviennent alors en fait un problème pour la majorité des techniques de génotypage SNP à haut débit. En effet, la plupart de ces techniques sont basées sur l'extension d'amorces. Si le site polymorphe d'intérêt n'est pas isolé d'autres sites polymorphes, ce type de techniques ne peut être utilisé. Les SNP sont donc très intéressants à développer car ils permettent la localisation de gènes mais la nature du génome de l'huître se prête mal au génotypage à haut débit.

La construction d'une carte génétique non plus simplement basée sur des marqueurs « anonymes » (microsatellites et AFLP) mais des gènes va permettre la comparaison de l'organisation des génomes d'espèces phylogénétiquement, proches ce qui apportera des éléments d'information quant à l'évolution de ces espèces. Cette approche est d'autant plus intéressante que d'autres cartes de liaison ont été développés au sein du genre *Crassostrea* et plus particulièrement chez *C. virginica* (Yu et al., 2003).

6.2 Détection de QTL

Le principal résultat de nos travaux est la primo-localisation de cinq QTL dont la part de variance expliquée pour les caractères survie et résistance à l'infection par OsHV1 est d'environ 50%. Cependant, les intervalles de confiance de ces régions chromosomiques sont encore larges et ne permettent pas de les localiser avec précision. Pour permettre la cartographie fine de ces régions, davantage de marqueurs moléculaires (SNP ou microsatellites) doivent être positionnés sur la carte de liaison consensus afin de réduire, l'intervalle moyen entre deux marqueurs, à une valeur comprise entre 5 et 10cM. Les QTL pourront alors être localisés de manière précise. La réduction de l'intervalle de confiance dans lequel se trouve un QTL est une étape essentielle vers l'identification des gènes impliqués dans la variance des caractères étudiés, un intervalle réduit contenant un plus petit nombre de gènes. La réduction de l'intervalle de confiance servira également à (1) séquencer ces régions à partir des clones des banques BAC disponibles chez *C. gigas* (Cunningham et al., 2006) pour tenter d'identifier la/les mutation(s) causale(s) ou QTN (Quantitative Trait Nucleotide) et (2) d'entreprendre la construction de cartes physiques à partir de la carte de liaison.

Des travaux de détection de QTL liés à l'hétérosis chez *C. gigas* (Hedgecock et al., 2007), lié à la résistance à des organismes pathogènes (MSX/Dermo) chez *C. virginica*, au

stade adulte (animaux de 2 à 3 ans ; Yu et Li, 2006), sont également en cours aux USA. Il sera alors intéressant de comparer la localisation de QTL liés à des caractères similaires (résistance à un pathogène) entre espèces pour déterminer si l'architecture génétique de ces caractères est de même nature.

Une seconde approche peut être envisagée pour localiser plus précisément les QTL. La cartographie fine de régions QTL en utilisant le déséquilibre de liaison et les études d'association est une alternative à la cartographie de QTL basée sur l'analyse de la ségrégation allélique dans une descendance (Goddard et al., 2005). Le modèle du déséquilibre de liaison utilise la non indépendance des allèles en deux lieux du génome. Dans une population isolée dont les croisements sont aléatoires, l'étendue du déséquilibre de liaison, après un événement de mutation, qui crée un déséquilibre avec tout le chromosome, décroît en espérance exponentiellement en fonction du nombre de générations et de la distance avec la mutation. Si la mutation est assez ancienne, seuls les marqueurs très proches sont en déséquilibre de liaison avec le lieu de la mutation. Ce principe permet d'envisager de localiser plus finement les gènes causaux si l'on dispose d'une carte dense de marqueurs tels que les SNP.

A long terme, la cartographie fine de QTL permettra de se diriger vers un schéma de sélection assistée par marqueurs (SAM) pour des caractères telle que la résistance au syndrome de mortalité estivale ou d'autres caractères d'intérêt aquacole ou commercial. L'intérêt d'une telle approche reste cependant à valider (rapport coût/bénéfice) du fait de l'héritabilité élevée du caractère.

Enfin, il apparaît évident et nécessaire de combiner ces approches de sélection (QTL, QTN et SAM) avec les manipulations du niveau de ploïdie qui sont réalisées chez l'huître creuse. La fixation de caractères d'intérêt, chez des individus triploïdes permettrait de proposer un produit très intéressant car il réunirait de nombreux avantages pour sa production (gamétogenèse réduite, vitesse de croissance) et sa commercialisation tout au long de l'année (animaux non laiteux).

6.3 Interactions hôte-pathogène

A l'issue du programme Morest, une interrogation demeure quant à la connaissance de la dynamique des événements de mortalités associés à des agents pathogènes, notamment des bactéries du genre *Vibrio* et le virus OsHV1, puisque aucun suivi serré n'avait pu être réalisé jusqu'à présent. Notre expérience de suivi quotidien a donc apporté un éclairage

nouveau sur cette dynamique et a révélé une forte corrélation entre la charge virale en OsHV1 et le pourcentage de mortalité cumulée chez nos familles F2. De plus, l'expérience de phénotypage des familles F2 a montré une réponse différentielle entre familles vis-à-vis de la charge virale et de la mortalité observée. L'hypothèse de la transmission inter-générationnelle de la résistance à la mortalité, précédemment proposée (Arzul et al., 2001; Barbosa-Solomieu et al., 2005; Boudry et al., 2008), est renforcée par la mise en évidence de QTL liés à la charge virale OsHV1. Cela suppose en fait une base génétique liée à la résistance à l'infection par le virus OsHV1 à partir de laquelle une sélection pour ce caractère aurait été réalisée au cours du projet « Morest ». Le suivi quotidien de la charge virale d'individus des lignées 'R' et 'S' au cours de mortalités estivales pourrait apporter des réponses à cette question.

En parallèle, les travaux de mise en évidence de gènes différentiellement exprimés au cours d'une réponse déclenchée par une infection virale associée à un événement de mortalité entre deux groupes de familles ou d'individus, doivent être poursuivis, notamment en ciblant des organes particuliers. Cela constitue une piste prometteuse pour la compréhension des mécanismes de cette réponse et des interactions complexes entre l'hôte et le pathogène. A terme, une expérience de phénotypage très similaire à celle réalisée au cours de l'été 2006, devrait être répétée pour combiner des données de génotypage et des données d'expression obtenues sur un large panel de gènes impliqués notamment dans la réponse à l'infection au virus OsHV1, pour rechercher des QTL d'expression (ou eQTL) liés à la variance pour le caractère de résistance à ce virus et ainsi mieux comprendre l'architecture génétique de ce caractère.

6.4 Applications pour la filière conchylicole

La poursuite des travaux de recherche engagés chez l'huître creuse, notamment en amélioration génétique, pourront déboucher vers des réalisations concrètes pour le secteur productif, telle que la sélection pour la résistance à des agents pathogènes comme l'herpès virus. Le nombre de publications sur le développement et, dans une moindre mesure l'application des marqueurs moléculaires en sélection a fortement augmenté au cours de la dernière décennie. Cependant, le transfert des résultats de publications vers des applications pratiques nécessite la résolution d'un grand nombre de contraintes techniques et génétiques qui sont rarement abordés dans la littérature. Le taux de réussite est susceptible d'augmenter en raison du développement de marqueurs de type SNP, de l'efficacité accrue des modèles de cartographie des QTL, et de la baisse des coûts de génotypage. Toutefois, certaines questions

fondamentales restent à résoudre, notamment en matière de caractères complexes, avant que la sélection assistée par marqueurs ne puisse être envisagée : on peut citer le développement de phénotypage à haut débit et de systèmes de précision pour la cartographie QTL, une meilleure compréhension des interactions génotype - environnement et le développement d'outils informatiques adaptés aux besoins des programmes de sélection moléculaire (Xu et al., 2008).

Actuellement, la filière conchylicole française est un secteur en pleine mutation car la demande croissante en fruits de mer, incluant les bivalves, croît et pour y faire face, la production devra augmenter (Helm et al., 2006). L'augmentation significative de l'approvisionnement à partir de la pêche traditionnelle des bivalves est peu probable puisque la plupart des stocks naturels sont surexploités ou à la limite maximale de leur exploitation durable. Cette augmentation proviendra donc probablement de l'aquaculture. En effet, le but actuel de plusieurs opérations est de restaurer les populations au niveau initial avant leur surexploitation. Les opérations futures de culture doivent être aussi efficaces que possible, non seulement pour des raisons de viabilité économique, mais aussi pour rendre optimale l'utilisation des aires de production qui se réduisent à cause de la pression démographique qui ne cesse d'augmenter et qui souffrent des activités humaines.

Toute augmentation de production implique une hausse en approvisionnement de naissain qui doit être fiable, abondant et peu coûteux. La collecte de juvéniles à partir des gisements naturels continuera à être significative mais ces aires sont limitées. L'apport essentiel en naissain en hausse viendra des écloseries. Il y a d'autres avantages pour la production du naissain d'écloserie par rapport à sa collecte en milieu naturel en plus de la fiabilité, comme la possibilité de produire une quantité qui peut faire face à la demande et la possibilité de produire des souches sélectionnées (Helm et al., 2006).

La filière conchylicole repose de plus en plus sur la capacité de ces écloseries à produire du naissain en quantité importante. Néanmoins, le risque associé est d'ordre sanitaire, avec l'émergence possible de nouveaux pathogènes.

Le programme MOREST avait notamment pour but d'étudier les bases génétiques de la résistance aux mortalités estivales. Suite à ce programme, le développement de programmes de sélection pour ce caractère est en discussion auxquels les résultats de nos travaux pourront contribuer. A terme, l'obtention de lignées résistantes à divers pathogènes,

viraux (OsHV1) ou bactériens (*Vibrio*), notamment en phase larvaire, est une perspective très intéressante pour la filière conchylicole car cela permettrait de diminuer l'incidence de ces agents pathogènes et ainsi d'améliorer les rendements de production. Ensuite, de caractères de qualité, tels que la forme de la coquille ou le poids de chair des individus sont des caractères qui pourraient entrer dans un schéma de sélection.

Enfin, depuis quelques années, la culture de nouvelles espèces, telles que les crevettes tropicales (genre *Penaeus*) ou l'ormeau (*Haliotis tuberculata*) émerge. Ces deux espèces sont également confrontées à des problèmes de mortalités anormales (syndrome viral « White spot » chez la crevette (Escobedo-Bonilla et al., 2008) et mortalités anormales associées à l'herpès virus ou à des bactéries du genre *Vibrio* chez l'ormeau (Hine, 1996). L'application des connaissances acquises chez l'huître creuse doit être réalisée dans de tels cas afin de proposer des solutions à ces filières de production.

REFERENCES

BIBLIOGRAPHIQUES

- AMCS Bulletin (Australian Marine Conservation Society). Marine and Coastal Updated. Bulletin Volume 20, No.2 Autumn 1998
- Ahmadian, A., B. Gharizadeh, A. C. Gustafsson, F. Sterky, P. Nyren, M. Uhlen et J. Lundeberg (2000). Single-Nucleotide Polymorphism analysis by pyrosequencing. *Analytical Biochemistry* 280(1): 103-110.
- Ahmed, M. et A. K. Sparks (1967). A preliminary study of chromosomes of two species of oysters (*Ostrea lurida* and *Crassostrea gigas*). *Journal of the Fisheries Research Board of Canada*. 24: 2155-2159.
- Alfonso, L. et C. S. Haley (1998). Power of different F2 schemes for QTL detection in livestock. *Animal Science* 66: 1-8.
- Allard, R. (1956). Formulas and tables to facilitate the calculation of recombination values in heredity. *Hilgardia* 24: 235-278.
- Allen, S. K. J. et S. L. Downing (1986). Performance of triploid Pacific oysters, *Crassostrea gigas* (Thunberg) I. Survival, growth, glycogen content, and sexual maturation in yearlings. *Journal of Experimental Marine Biology and Ecology* 102(2-3): 197-208.
- Anderson, E. C. et J. C. Garza (2006). The Power of Single-Nucleotide polymorphisms for large-scale parentage inference. *Genetics* 172(4): 2567-2582.
- Aranishi, F. et T. Okimoto (2004). Genetic relationship between cultured populations of Pacific oyster revealed by RAPD analysis. *Journal of Applied Genetics* 45(4): 435-443.
- Arzul, I., T. Renault et C. Lipart (2001). Experimental herpes-like viral infections in marine bivalves: demonstration of interspecies transmission. *Disease Aquatic Organisms* 46(1): 1-6.
- Baranski, M., S. Loughnan, C. M. Austin et N. Robinson (2006). A microsatellite linkage map of the blacklip abalone, *Haliotis rubra*. *Animal Genetics* 37(6): 563-570.
- Baranski, M., N. Robinson, S. Loughnan, M. Rourke, B. Hayes et C. Austin (2007). Detection of QTL for growth rate in abalone using selective DNA pooling. *Aquaculture* 272 (Supplement 1): S242-S243.
- Barbosa-Solomieu, V., L. Degremont, R. Vazquez-Juarez, F. Ascencio-Valle, P. Boudry et T. Renault (2005). Ostreid Herpesvirus 1 (OsHV-1) detection among three successive generations of Pacific oysters, *Crassostrea gigas*. *Virus Research* 107(1): 47-56.
- Basten, C., B. S. Weir et Z. B Zeng (1997). QTL Cartographer: a reference manual and tutorial for QTL mapping. Raleigh, N.C.: Department of Statistics, North Carolina State University.
- Bazin, E., L. Duret, S. Penel et N. Galtier (2005). Polymorphix: a sequence polymorphism database. *Nucleic Acids Research*. 33 (supplement1): 481-484.

- Beattie, J., K. Chew et W. Hershberger (1980). Differential survival of selected strains of Pacific oysters (*Crassostrea gigas*) during summer mortality. *Proceedings of National Shellfish Association* 70: 184-189.
- Beaumont, A. R., C. M. Beveridge et M. D. Budd (1983). Selection and heterozygosity within single families of the mussel *Mytilus edulis* L. *Marine Biology Letters* 4: 151-161.
- Beavis, W. et D. Grant (1991). A linkage map based on information from F2 populations of maize (*Zea mays* L.). *Theoretical Applied Genetics* 88: 636-644.
- Beavis, W. (1998). QTL Analyses: power, precision and accuracy In: *Molecular Dissection of Complex Trait*. New York.
- Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste et F. Bonhomme (1996). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations, Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).
- Bierne, N., S. Launey, Y. Naciri-Graven et F. Bonhomme (1998). Early effect of inbreeding as revealed by microsatellite analyses on *Ostrea edulis* larvae. *Genetics* 148(4): 1893-1906.
- Bierne, N., A. Tsitrone et P. David (2000). An inbreeding model of associative overdominance during a population bottleneck. *Genetics* 155(4): 1981-1990.
- Borts, R. et J. Haber (1987). Meiotic recombination in yeast: alteration by multiple heterozygosities. *Science* 237: 1459-1465.
- Boudry, P., S. Heurtebise, B. Collet, F. Cornette et A. Gérard (1998). Differentiation between populations of the Portuguese oyster, *Crassostrea angulata* (Lamarck) and the Pacific oyster, *Crassostrea gigas* (Thunberg), revealed by mtDNA RFLP analysis. *Journal of Experimental Marine Biology and Ecology* 226: 279-291.
- Boudry, P., B. Collet, F. Cornette, V. Hervouet et F. Bonhomme (2002). High variance in reproductive success of the Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. *Aquaculture* 204(3-4): 283-296.
- Boudry, P., S. Heurtebise et S. Lapègue (2003a). Mitochondrial and nuclear DNA sequence variation of presumed *Crassostrea gigas* and *Crassostrea angulata* specimens : a new oyster species in Hong Kong ? *Aquaculture* 228: 15-25.
- Boudry, P. et L. Dégremont (2008). Etude des bases génétiques de la mortalité estivale du naissain et des possibilités d'amélioration de la survie par sélection. *Mortalités estivales de l'huître creuse Crassostrea gigas: Défi MOREST*. H. McCombie, Jean-François Samain, Edition Quae.
- Bouilly, K., A. Leitao, H. McCombie et S. Lapegue (2003). Impact of atrazine on aneuploidy in pacific oysters, *Crassostrea gigas*. *Environmental Toxicology and Chemistry* 22(1): 219-23.

- Bouilly, K., A. Leitaó, R. Chaves, H. Guedes-Pinto, P. Boudry et S. Lapegue (2005). Endonuclease banding reveals that atrazine-induced aneuploidy resembles spontaneous chromosome loss in *Crassostrea gigas*. *Genome* 48(1): 177-80.
- Bouilly, K., B. Gagnaire, M. Bonnard, H. Thomas-Guyon, T. Renault, P. Miramand et S. Lapegue (2006). Effects of cadmium on aneuploidy and hemocyte parameters in the Pacific oyster, *Crassostrea gigas*. *Aquatic Toxicology* 78(2): 149-56.
- Bouilly, K., M. Bonnard, B. Gagnaire, T. Renault et S. Lapegue (2007). Impact of diuron on aneuploidy and hemocyte parameters in Pacific oyster, *Crassostrea gigas*. *Archives of Environmental Contamination and Toxicology* 52(1): 58-63.
- Bovenhuis, H. et R. Spelman (2000). Selective genotyping to detect quantitative trait loci for multiple traits in outbred populations. *Journal of Dairy Science* 83: 173-180.
- Boyd, M. J. (2003). Nonindigenous marine species in Humboldt Bay, California. In Abstracts: Third International Conference on Marine Bioinvasions March 16-19, 2003, Scripps Institution of Oceanography La Jolla, California.
- Britten, H. (1996). Meta-analysis of the association between multilocus heterozygosity and fitness. *Evolution* 50: 2158-2164.
- Broman, K. W., J. C. Murray, V. C. Sheffield, R. L. White et J. L. Weber (1998). Comprehensive human genetic maps: individual and sex-specific variation in recombination. *American Journal of Human Genetic* 63(3): 861-9.
- Brookes, A. J. (1999). The essence of SNPs. *Gene* 234(2): 177-86.
- Brooks, W. K. (1879). Abstract of observations upon fertilization of oyster eggs and embryology of American oyster. *American Journal of Sciences* 18: 425-427.
- Brown, J. et V. Sundaresan (1991). A recombination hotspot in the maize A1 intragenic region. *Theoretical Applied Genetics* 81: 185-188.
- Buckleton, J. et C. Triggs (2006). Dealing with allelic dropout when reporting the evidential value in DNA relatedness analysis. *Forensic Science International* 160 (2-3): 134-9.
- Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129(3): 897-907.
- Buroker, N. E., W. K. Hershberger et K. K. Chew (1975). Genetic variation in the Pacific Oyster *Crassostrea gigas* *Journal of the Fisheries Research Board of Canada*. 32(2471-2477).
- Buroker, N. E., W. Hershberger et K. Chew (1979). Population genetics of the family Ostreidae : I. Intraspecific studies of *Crassostrea gigas* and *Saccostrea commercialis*. *Marine Biology* 54: 157-169.
- Buroker, N. E. (1983). Sexuality with respect to shell length and group size in the Japanese oyster *Crassostrea gigas* *Malacologia* 23: 271-279.

- Bustin, S. A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology* 25(2): 169-93.
- Callen, D. F., A. D. Thompson, Y. Shen, H. A. Phillips, R. I. Richards, J. C. Mulley et G. R. Sutherland (1993). Incidence and origin of "null" alleles in the (AC)_n microsatellite markers. *American Journal of Human Genetic* 52(5): 922-927.
- Carter, T. C. et D. S. Falconer (1951). Stocks for detecting linkage in the mouse, and the theory of their design. *Journal of Genetics* 5: 307-323.
- Castano-Sanche, C., K. Fuji, A. Ozaki, O. Hasegawa, T. Sakamoto, K. Morishima, I. Nakayama, A. Fujiwara, T. Masaoka, H. Okamoto, M. Okauti, K. Hayashida, M. Tagami, J. Kawai, Y. Hayashizaki et N. Okamoto (2007). High-density linkage map of the Japanese flounder, *Paralichthys olivaceus*. *Aquaculture* 272 (Supplement 1): S248.
- Causse, M., J. Rocher, A. Henry, A. Charcosset, J. Prioul et D. de Vienne (1995). Genetic dissection of the relationship between carbon metabolism and early growth in maize, with emphasis on key-enzyme loci. *Molecular Breeding* 3: 259-272.
- Chakravarti, A., L. K. Lasher et J. E. Reefer (1991). A Maximum likelihood method for estimating genome length using genetic linkage data. *Genetics* 128(1): 175-182.
- Chatziplis, D., C. Batargias, C. S. Tsigenopoulos, A. Magoulas, S. Kollias, G. Kotoulas, F. A. M. Volckaert et C. S. Haley (2007). Mapping quantitative trait loci in European sea bass (*Dicentrarchus labrax*): The BASSMAP pilot study. *Aquaculture* 272 (Supplement 1): S172-S182.
- Chistiakov, D. A., B. Hellemans, C. S. Haley, A. S. Law, C. S. Tsigenopoulos, G. Kotoulas, D. Bertotto, A. Libertini et F. A. M. Volckaert (2005). A Microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L. *Genetics* 170(4): 1821-1826.
- Chistiakov, D. A., B. Hellemans et F. A. M. Volckaert (2006) Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture* 255: 1-29.
- Cho, R. J., M. Mindrinos, D. R. Richards, R. J. Sapolsky, M. Anderson, E. Drenkard, J. Dewdney, T. L. Reuber, M. Stammers, N. Federspiel, A. Theologis, W. H. Yang, E. Hubbell, M. Au, E. Y. Chung, D. Lashkari, B. Lemieux, C. Dean, R. J. Lipshutz, F. M. Ausubel, R. W. Davis et P. J. Oefner (1999). Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nature Genetics* 23(2): 203-207.
- Churchill, G. A. et R. W. Doerge (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138(3): 963-71.
- Clop, A., F. Marcq, H. Takeda, D. Pirottin, X. Tordoir, B. Bibe, J. Bouix, F. Caiment, J. M. Elsen, F. Eychenne, C. Larzul, E. Laville, F. Meish, D. Milenkovic, J. Tobin, C. Charlier et M. Georges (2006). A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics* 38(7): 813-818.

- Coimbra, M. R. M., K. Kobayashi, S. Koretsugu, O. Hasegawa, E. Ohara, A. Ozaki, T. Sakamoto, K. Naruse et N. Okamoto (2003). A genetic linkage map of the Japanese flounder, *Paralichthys olivaceus*. *Aquaculture* 220(1-4): 203-218.
- Collins, A., J. Teague, B. J. Keats et N. E. Morton (1996). Linkage Map Integration. *Genomics* 36(1): 157-162.
- Collins, F. S., L. D. Brooks et A. Chakravarti (1998). A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research* 8(12): 1229-31.
- Comps, M. et J. L. Duthoit (1976). Infection virale associée à la maladie des branchies de l'huître creuse *Crassostrea angulata*. *Comptes rendus de l'Académie des Sciences, Paris* 283: 1595-1597.
- Conde, L., J. M. Vaquerizas, H. Dopazo, L. Arbiza, J. Reumers, F. Rousseau, J. Schymkowitz et J. Dopazo (2006). PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Research*. 34(supplement 2): 621-625.
- Coquoz, R. (2003). Preuve par l'ADN. La génétique au service de la justice, Collection sciences forensiques.
- Cormier, R. T., K. H. Hong, R. B. Halberg, T. L. Hawkins, P. Richardson, R. Mulherkar, W. F. Dove et E. S. Lander (1997). Secretory phospholipase Pla2g2a confers resistance to intestinal tumorigenesis. *Nature Genetics* 17(1): 88-91.
- CSIRO (2002). Australians select best Pacifics collaborative breeding programs *Fish Farming International* 29(10): 34-35.
- Cunningham, C., J. Hikima, M. J. Jenny, R. W. Chapman, G. C. Fang, C. Saski, M. L. Lundqvist, R. A. Wing, P. M. Cupit, P. S. Gross, G. W. Warr et J. P. Tomkins (2006). New resources for marine genomics: bacterial artificial chromosome libraries for the Eastern and Pacific oysters (*Crassostrea virginica* and *C. gigas*). *Marine Biotechnology* (NY) 8(5): 521-33.
- Curole, J. P. et D. Hedgecock (2005). High Frequency of SNPs in the Pacific Oyster genome. http://intl-pag.org/13/abstracts/PAG13_W026.html. Plant and Animal Genome XIII Symposium.
- Cutter, A. D., S. E. Baird et D. Charlesworth (2006). High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. *Genetics* 174(2): 901-13.
- Danzmann, R. G., T. R. Jackson et M. M. Ferguson (1999). Epistasis in allelic expression at upper temperature tolerance QTL in rainbow trout. *Aquaculture* 173(1-4): 45-58.
- Darvasi, A. et M. Soller (1992). Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theoretical and Applied Genetics* 85(2): 353-359.

- Darvasi, A., A. Weinreb, V. Minke, J. I. Weller et M. Soller (1993). Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134(3): 943-951.
- Darvasi, A. (1997). The effect of selective genotyping on QTL mapping accuracy. *Mammalian Genome* 8(1): 67-68.
- Darvasi, A. (1998). Experimental strategies for the genetic dissection of complex traits in animal models. *Nature Genetic* 18(1): 19-24.
- David, E., P. Boudry, L. Dégremont, A. Tanguy, N. Quéré, J. F. Samain et D. Moraga (2007). Genetic polymorphism of glutamine synthetase and delta-9 desaturase in families of Pacific oyster *Crassostrea gigas* and susceptibility to summer mortality. *Journal of Experimental Marine Biology and Ecology* 349(2): 272-283.
- David, P. (1998). Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity* 80: 531-537.
- de Koning, D. J., P. M. Visscher, S. A. Knott et C. S. Haley (1998). Strategies for QTL detection in half-sib populations. *Animal Science* 67: 257-268.
- De la Vega, F. M., K. D. Lazaruk, M. D. Rhodes et M. H. Wenz (2005). Assessment of two flexible and compatible SNP genotyping platforms: TaqMan SNP genotyping assays and the SNPlex genotyping system. *Mutation Research* 573(1-2): 111-135.
- de Vicente, M. C. et S. D. Tanksley (1993). QTL analysis of transgressive segregation in an interspecific tomato cross. *Genetics* 134: 585-596.
- de Vienne, D. (1998). Les marqueurs moléculaires en génétique et biotechnologies végétales, INRA éditions.
- Dégremont, L. (2003). Etude des bases génétiques de la mortalité estivale et des relations avec la croissance chez les juvéniles de l'huître creuse *Crassostrea gigas*., Thèse de doctorat. Université de Caen - Basse Normandie: 333p.
- Dégremont, L., E. Bedier, P. Soletchnik, M. Ropert, A. Huvet, J. Moal, J.-F. Samain et P. Boudry (2005). Relative importance of family, site, and field placement timing on survival, growth, and yield of hatchery-produced Pacific oyster spat (*Crassostrea gigas*). *Aquaculture* 249: 213-229.
- Dégremont, L., B. Ernande, E. Bedier et P. Boudry (2007). Summer mortality of hatchery-produced Pacific oyster spat (*Crassostrea gigas*). I. Estimation of genetic parameters for survival and growth. *Aquaculture* 262(1): 41-53.
- Derayat, A., R. D. Houston, D. R. Guy, A. Hamilton, J. Ralph, N. Spreckley, J. B. Taggart, B. J. McAndrew et C. S. Haley (2007). Mapping QTL affecting body lipid percentage in Atlantic salmon (*Salmo salar*). *Aquaculture* 247: 250-251.

- Devey, M. E., T. A. Fiddler, B. H. Liu, S. J. Knapp et B. D. Neale (1994). An RFLP linkage map for loblolly pine based on a three-generation outbred pedigree. *Theoretical Applied Genetics* 88: 273-278.
- Doerge, R. W., Z. Zeng et B. Weir (1997). Statistical issues in the search for genes affecting quantitative traits in experimental populations. *Statistical Sciences* 12: 195-219.
- Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nature Review Genetics* 3(1): 43-52.
- Douches, D. et C. Quiros (1988). Genetic recombination in a diploid synaptic mutant and a *Solanum tuberosum S.chacoense* diploid hybrid. *Heredity* 60: 183-191.
- Drew, R. E., H. Schwabl, P. A. Wheeler et G. H. Thorgaard (2007). Detection of QTL influencing cortisol levels in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 272(1): 183-194.
- Duret, L. et D. Mouchiroud (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proceedings of National Academic Sciences USA* 96(8): 4482-4487.
- Durham, R. E., P. C. Liou, F. G. Gmitter et G. A. Moore (1992). Linkage of restriction fragment length polymorphisms and isozymes in *Citrus*. *Theoretical Applied Genetics* 84: 39-48.
- Edwards, M. D., C. W. Stuber et J. F. Wendel (1987). Molecular marker-facilitated investigations of quantitative-traits loci in maize. I. Numbers, genomic distribution and types of gene action. *Genetics* 116: 113-125.
- Elsen, J. M., B. Mangin, B. Goffinet et C. Chevalet (1994). Optimal structure of protocol design for building genetic linkage maps in livestock. *Theoretical Applied Genetics* 88: 129-134.
- Erickson, D. (2005). Quantitative trait loci: mapping the future of QTL's. *Heredity* 95(6): 417-8.
- Ernande, B. (2001). Développement, plasticité phénotypique et évolution, Université de La Rochelle: 274p.
- Ernande, B., P. Boudry, J. Clobert et J. Haure (2004). Plasticity in resource allocation based on life history traits in the Pacific oyster, *Crassostrea gigas*. I. Spatial variation in food abundance. *Journal of Evolutionary Biology* 17(2): 342-56.
- Escobedo-Bonilla, C. M., V. Alday-Sanz, M. Wille, P. Sorgeloos, M. B. Pensaert et H. J. Nauwynck (2008). A review on the morphology, molecular characterization, morphogenesis and pathogenesis of white spot syndrome virus. *Journal of Fish Diseases* 31(1): 1-18.

- Eshed, Y., M. Abu-Abied, Y. Saranga et D. Zamir (1992). *Lycopersicon esculentum* lines containing small overlapping introgressions from *L. pennellii*. *Theoretical Applied Genetics* 83: 1027-1034.
- Evans, S., S. E. Matson, J. Brake et C. Langdon (2004). The effects of inbreeding on performance traits of adult Pacific oysters (*Crassostrea gigas*). *Aquaculture* 230: 89-98.
- Fabioux, C. (2004). Origine et développement des cellules germinales de l'huître creuse *Crassostrea gigas* : Intérêt pour le contrôle de la reproduction en éclosion. Thèse de doctorat de l'Université de Bretagne occidentale. Brest: 219p.
- Falconer, D. S. et T. F. C. Mackay (1996). *Introduction to Quantitative Genetics*. Essex, England., Longman Group Ltd.
- FAO (2007). *Yearbook of Fisheries statistics: Summary Tables*. World aquaculture production of fish, crustaceans, molluscs, etc by principal species in 2005. <http://www.fao.org/fi/statist/summtab/default.asp>.
- Faure, S., J. Noyer, J. Horry, F. Bakry, C. Lanaud et D. González de León (1993). A molecular marker based linkage map of diploid bananas. *Theoretical and Applied Genetics* 87: 517-526.
- Ferreira, M. E., P. H. Williams et T. C. Osborn (1994). RFLP mapping of *Brassica napus* using doubled haploid lines. *Theoretical Applied Genetics* 89: 615-621.
- Fleury, E., C. Fabioux, C. Lelong, P. Favrel et A. Huvet (2008). Characterization of a gonad-specific transforming growth factor beta superfamily member differentially expressed during the reproductive cycle of the oyster *Crassostrea gigas*. *Gene* 410: 187-196.
- Fleury, P. G., F. Cornette, H. Claude, H. Palvadeau, S. Robert, P. d'Amico, P. Le Gall et C. Vercelli (2001). Réseau d'évaluation de la croissance de l'huître creuse sur les côtes françaises. REMORA. Résultats des stations nationales. Année 2000. Rapport interne Ifremer DRV/RA/RST/01-02.
- Fotherby, H. A., H. K. Mogahadam, R. G. Danzmann et M. M. Ferguson (2007). Detection of quantitative trait loci for body weight, condition factor and age at sexual maturation in North American Atlantic salmon (*Salmo salar*) and comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). *Aquaculture* 247: 256-257.
- Franch, R., B. Louro, M. Tsalavouta, D. Chatziplis, C. S. Tsigenopoulos, E. Sarropoulou, J. Antonello, A. Magoulas, C. C. Mylonas, M. Babbucci, T. Patarnello, D. M. Power, G. Kotoulas et L. Bargelloni (2006). A genetic linkage map of the hermaphrodite teleost fish *Sparus aurata* L. *Genetics* 174(2): 851-861.
- Fridman, E., F. Carrari, Y. S. Liu, A. R. Fernie et D. Zamir (2004). Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* 305(5691): 1786-9.

- Friedman, C. S., R. M. Estes, N. A. Stokes, C. A. Burge, J. S. Hargove, B. J. Barber, R. A. Elston, E. M. Burrenson et K. S. Reece (2005). Herpes virus in juvenile Pacific oysters *Crassostrea gigas* from Tomales Bay, California, coincides with summer mortality episodes. *Disease Aquatic Organisms* 63(1): 33-41.
- Fries, R. et G. Durstewitz (2001). Digital DNA signatures for animal tagging. *Nature Biotechnology* 19(6): 508.
- Gaffney, P. M. et S. K. J. Allen (1993). Hybridization among *Crassostrea* species : a review. *Aquaculture* 116: 1-13.
- Garcia-Dorado, A. et A. Gallego (1992). On the use of the classical tests for detecting linkage. *Journal of Heredity* 83: 143-146.
- Gaut, B. S. et M. T. Clegg (1993). Nucleotide polymorphism in the Adh1 locus of pearl millet (*Pennisetum glaucum*) (Poaceae). *Genetics* 135(4): 1091-7.
- Gebhardt, C., E. Ritter, A. Barone, D. T. B. Walkemeier, U. Schachtschabel, H. Kaufmann, R. Thompson, M. W. Bonierbale, M. Ganal, S. D. Tanksley et F. Salamini (1991). RFLP maps of potato and their alignment with the homeologous tomato genome. *Theoretical Applied Genetics* 83: 49-57.
- Gérard, A., Y. Naciry-Graven, P. Boudry, S. Launey, S. Heurtebise, C. Ledu et P. Phelipot (1995). Contrôle de la gamétogenèse des huîtres creuses et plates. Relations "reproduction" et "génétique". Groupe de Travail sur la reproduction des Mollusques (Nantes 14-15 Novembre): 99-111.
- Gharbi, K., A. Gautier, R. G. Danzmann, S. Gharbi, T. Sakamoto, B. Hoyheim, J. B. Taggart, M. Cairney, R. Powell, F. Krieg, N. Okamoto, M. M. Ferguson, L.-E. Holm et R. Guyomard (2006). A Linkage Map for Brown Trout (*Salmo trutta*): Chromosome Homeologies and Comparative Genome Organization With Other Salmonid Fish. *Genetics* 172(4): 2405-2419.
- Gilbey, J., E. Verspoor, T. A. Mo, E. Sterud, K. Olstad, S. Hytterod, C. Jones et L. Noble (2006). Identification of genetic markers associated with *Gyrodactylus salaris* resistance in Atlantic salmon *Salmo salar*. *Diseases of Aquatic Organisms* 71(2): 119-129.
- Gjedem, T. (1985). Improvement of productivity through breeding schemes. *Geojournal* 10: 233-241.
- Glazier, A. M., J. H. Nadeau et T. J. Aitman (2002). Findings genes that underlie complex traits. *Science* 298: 2345-2349.
- Goddard, M. E. et T. H. E. Meuwissen (2005). The use of linkage disequilibrium to map quantitative trait loci. *Australian Journal of Experimental Agriculture* 45: 837-845.
- Goldstein, D. B., G. W. Roemer, D. A. Smith, D. E. Reich, A. Bergman et R. K. Wayne (1999). The use of microsatellite variation to infer population structure and demographic history in a natural model system. *Genetics* 151(2): 797-801.

- Gonzalez-Tizon, A. M., A. Martinez-Lage, I. Rego, J. Ausio et J. Mendez (2000). DNA content, karyotypes, and chromosomal location of 18S-5.8S-28S ribosomal loci in some species of bivalve molluscs from the Pacific Canadian coast. *Genome* 43(6): 1065-72.
- Gouletquer, P. (1997). Natural reproductive cycle of *Crassostrea gigas* Proceedings: The natural and controlled reproduction of cultivated bivalves in France: symposium report, Ifremer: 7-19.
- Gouletquer, P., M. Wolowicz, A. Latala, P. Geairon, A. Huvet et P. Boudry (1999). Comparative analysis of oxygen consumption rates between cupped oyster spat of *Crassostrea gigas* of French, Japanese, Spanish and Taiwanese origins. *Aquatic Living Resources* 12: 271-277.
- Gouletquer, P. (2005). CABI Aquaculture Compendium. Chapter Culture aquatic animals species.
- Gradiner, J. M., E. H. Coe, S. Melia-Hancock, D. A. Hoisington et S. Chao (1993). Development of a core RFLP map in maize using an immortalized F2 population. *Genetics* 134: 917-930.
- Graner, A., A. Jahorr, J. Schondelmaier, H. Siedler, K. Pille'n, G. Fishbeck, G. Wenzel et R. Herrmann (1991). Construction of an RFLP map of barley. *Theoretical Applied Genetics* 83: 250-256.
- Grattapaglia, D. et R. Sederoff (1994). Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. *Genetics* 137: 1121-1137.
- Gregory, T. R. (2005). Animal Genome Size Database (<http://www.genomesize.com>).
- Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges et R. Snell (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research* 12(2): 222-31.
- Grivet, L., J. C. Glaszmann, M. Vincentz, F. da Silva et P. Arruda (2003). ESTs as a source for sequence polymorphism discovery in sugarcane: example of the Adh genes. *Theoretical and Applied Genetics* 106(2): 190-197.
- Grizel, H. et M. Héral (1991). Introduction into France of the Japanese oyster (*Crassostrea gigas*). *Journal du Conseil International pour Exploration de la Mer* 47: 399-403.
- Groenen, M. A., H. H. Cheng, N. Bumstead, B. F. Benkel, W. E. Briles, T. Burke, D. W. Burt, L. B. Crittenden, J. Dodgson, J. Hillel, S. Lamont, A. P. de Leon, M. Soller, H. Takahashi et A. Vignal (2000). A consensus linkage map of the chicken genome. *Genome Res* 10(1): 137-147.
- Groover, A., C. Williams, M. Devey, J. Lee et D. Neale (1995). Sex-related differences in meiotic recombination frequency in *Pinus taeda*. *Journal of Heredity* 86: 157-158.

- Gruber, J. D., P. B. Colligan et J. K. Wolford (2002). Estimation of single nucleotide polymorphism allele frequency in DNA pools by using pyrosequencing. *Human Genetics* 110(5): 395-401.
- Gueguen, Y., J.-P. Cadoret, D. Flament, C. Barreau-Roumiguere, A.-L. Girardot, J. Garnier, A. Hoareau, E. Bachere et J.-M. Escoubas (2003). Immune gene discovery by expressed sequence tags generated from hemocytes of the bacteria-challenged oyster, *Crassostrea gigas*. *Gene* 303: 139-145.
- Guo, X., D. Hedgecock, W. Hershberger, K. Cooper et S. K. J. Allen (1998). Genetic determinants of protandric sex in *Crassostrea* oyster. *Evolution* 52: 394-402.
- Gut, I. G. (2001). Automation in genotyping of single nucleotide polymorphisms. *Human Mutation* 17(6): 475-92.
- Gut, I. G. (2004). DNA analysis by MALDI-TOF mass spectrometry. *Human Mutation* 23(5): 437-441.
- Guyomard, R., S. Mauger, K. Tabet-Canale, S. Martineau, C. Genet, F. Krieg et E. Quillet (2006). A Type I and Type II microsatellite linkage map of Rainbow trout (*Oncorhynchus mykiss*) with presumptive coverage of all chromosome arms. *BMC Genomics* 7(1): 302.
- Haanstra, J. P., R. Lauge, F. Meijer-Dekens, G. Bonnema, P. J. de Wit et P. Lindhout (1999). The Cf-ECP2 gene is linked to, but not part of, the Cf-4/Cf-9 cluster on the short arm of chromosome 1 in tomato. *Molecular and General Genetics* 262(4-5): 839-845.
- Hackett, C. A. et L. B. Broadfoot (2003). Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* 90(1): 33-38.
- Haidle, L., J. Janssen, K. Gharbi, R. G. Danzmann et M. M. Ferguson (2007). Identifying regions of the genome responsible for early maturation and associated traits in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 247: 265-266.
- Haldane, J. B. S. (1919). The combination of linkage values and the calculation of distances between the loci of linked factors. *Journal of Genetics* 8: 229-309.
- Haley, C. S. et S. A. Knott (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69(4): 315-324.
- Haley, C. S., S. A. Knott et J. M. Elsen (1994). Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* 136(3): 1195-1207.
- Hall, T. (1997). Bioedit (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>).
- Harry, H. W. (1985). Synopsis of the supraspecific classification of living oysters (Bivalvia : Gryphaeida and Ostreidae). *Veliger* 28: 121-158.

- Haure, J., A. Huvet, H. Palvadeau, M. Nourry, C. Penisson, J. L. Martin et P. Boudry (2003). Feeding and respiratory time activities in the cupped oysters *Crassostrea gigas*, *Crassostrea angulata* and their hybrids. *Aquaculture* 218: 539-551.
- Hedgecock, D. et F. L. Sly (1990). Genetic drift and effective population sizes of hatched-propagated stocks of the Pacific oyster *Crassostrea gigas*. *Aquaculture* 88: 21-38.
- Hedgecock, D., V. Chow et R. S. Waples (1992). Effective population numbers of shellfish broodstocks estimated from temporal variance in allelic frequencies. *Aquaculture* 108: 215-232.
- Hedgecock, D., D. J. McGoldrick et B. L. Bayne (1995). Hybrid vigor in Pacific oysters: an experimental approach using crosses among inbred lines. *Aquaculture* 137: 285-298.
- Hedgecock, D., P. M. Gaffney, P. Gouletquer, X. Guo, K. Reece et W. Warr Gregory (2005). The case for sequencing the Pacific oyster genome. *Journal of Shellfish Research* 24(2): 429-441.
- Hedgecock, D., J.-Z. Lin, S. DeCola, C. D. Haudenschild, E. Meyer, D. T. Manahan et B. Bowen (2007). Transcriptomic analysis of growth heterosis in larval Pacific oysters (*Crassostrea gigas*). *Proceedings of National Academic of Sciences of the USA* 104(7): 2313-2318.
- Hedgecock, D., G. M. L. Perry et M. L. Voigt (2007). Mapping heterosis QTL in the Pacific oyster *Crassostrea gigas*. *Aquaculture* 272 (Supplement 1): S267-S268.
- Helm, M. M., N. Bourne et A. Lovatelli (2006). *Ecloserie de Bivalves. Un manuel Pratique*. Rome, FAO, un document techniques sur les pêches.
- Hershberger, W., J. A. Perdue et J. Beattie (1984). Genetic selection and systematic breeding in Pacific oyster culture. *Aquaculture* 39: 237-245.
- Hine, P. M. (1996). Southern hemisphere mollusc diseases and an overview of associated risk assessment problems. *Revue Sciences et Techniques* 1(2): 563-577.
- Holland, J. B. (2007). Genetic architecture of complex traits in plants. *Current Opinion in Plant Biology* 10(2): 156-161.
- Hopkins, C. (2001). A review of introductions and transfers of alien marine species in the North Sea Area. In Report to Directorate for Nature Management, DN, Oslo, Norway, 96 pp. <http://www.dirnat.no/archive/attachments/01/04/CHMDA071.pdf>
- Houston, R. D., C. S. Haley, A. Hamilton, D. R. Guy, A. E. Tinch, J. B. Taggart, B. J. McAndrew et S. C. Bishop (2008). Major Quantitative Trait Loci Affect Resistance to Infectious Pancreatic Necrosis in Atlantic Salmon (*Salmo salar*). *Genetics* 178(2): 1109-1115.
- Hubert, S. et D. Hedgecock (2004). Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. *Genetics* 168(1): 351-62.

- Hughes, J. R. (2002). *Crassostrea gigas*. Marine life Information Network: Biology and sensitivity key information sub-program., Plymouth: Marine Biological Association of the United Kingdom.
- Humphray, S., C. Scott, R. Clark, B. Marron, C. Bender, N. Camm, J. Davis, A. Jenks, A. Noon, M. Patel, H. Sehra, F. Yang, M. Rogatcheva, D. Milan, P. Chardon, G. Rohrer, D. Nonneman, P. de Jong, S. Meyers, A. Archibald, J. Beever, L. Schook et J. Rogers (2007). A high utility integrated map of the pig genome. *Genome Biology* 8(7): R139.
- Huvet, A., P. Boudry, M. Ohresser, C. Delsert et F. Bonhomme (2000). Variable microsatellites in the Pacific oyster *Crassostrea gigas* and other cupped oyster species. *Animal Genetics* 31(1): 71-72.
- Huvet, A., K. Balabaud, N. Bierne et P. Boudry (2001). Microsatellite analysis of 6-hour-old embryos reveals no preferential intraspecific fertilization between cupped oysters *Crassostrea gigas* and *Crassostrea angulata*. *Marine Biotechnology* 3: 448-453.
- Huvet, A., A. Gérard, C. Ledu, P. Phélipot, S. Heurtebise et P. Boudry (2002). Is fertility of hybrids enough to conclude that the two oysters *Crassostrea gigas* and *Crassostrea angulata* are the same species? *Aquatic Living research* 15: 15-52.
- Huvet, A., C. Fabioux, H. McCombie, S. Lapegue et P. Boudry (2004). Natural hybridization in genetically differentiated populations of *Crassostrea gigas* and *C. angulata* highlighted by sequence variation in flanking regions of a microsatellite locus. *Marine Ecology Progress Series* 272: 141-152.
- Huvet, A., A. Herpin, L. Degremont, Y. Labreuche, J.-F. Samain et C. Cunningham (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.
- Ibarra, A. M., P. Cruz et B. A. Romero (1995). Effects of inbreeding on growth and survival of self-fertilized catarina scallop larvae, *Argopecten circularis*. *Aquaculture* 134(1-2): 37-47.
- ICSEMS. (2000). 2007, from <http://www.ciesm.org/atlas/Crassostreagigas.html>.
- Ihrke, C. et W. Kronstad (1975). Genetic recombination in maize as affected by EDTA and DMSO. *Crop Science* 15: 429-431.
- Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution* 2: 13-34.
- Jackson, T. R., M. M. Ferguson, R. G. Danzmann, A. G. Fishback, P. E. Ihssen, M. O'Connell et T. J. Crease (1998). Identification of two QTL influencing upper temperature tolerance in three rainbow trout (*Oncorhynchus mykiss*) half-sib families. *Heredity* 80(2): 143-151.
- Jansen, R. C. (1993). Interval Mapping of Multiple Quantitative Trait Loci. *Genetics* 135(1): 205-211.

- Jarrel, D. C., M. L. Roose, S. N. Traugh et R. S. Kupper (1992). A genetic map of citrus based on the segregation of isozymes and RFLPs in an intergeneric cross. *Theoretical Applied Genetics* 84: 49-56.
- Jenny, M. J., R. W. Chapman, A. Mancina, Y. A. Chen, D. J. McKillen, H. Trent, P. Lang, J. M. Escoubas, E. Bachere, V. Boulo, Z. J. Liu, P. S. Gross, C. Cunningham, P. M. Cupit, A. Tanguy, X. Guo, D. Moraga, I. Boutet, A. Huvet, S. De Guise, J. S. Almeida et G. W. Warr (2007). A cDNA microarray for *Crassostrea virginica* and *C. gigas*. *Marine Biotechnology (NY)* 9(5): 577-91.
- Jiang, C. et Z. B. Zeng (1997). Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica* 101(1): 47-58.
- Ka-Shu Wong, G., B. Liu, J. Wang, Y. Zhang, X. Yang, Z. Zhang, Q. Meng, J. Zhou, D. Li, J. Zhang, P. Ni, S. Li, L. Ran, H. Li, J. Zhang, R. Li, S. Li, H. Zheng, W. Lin, G. Li, X. Wang, W. Zhao, J. Li, C. Ye, M. Dai, J. Ruan, Y. Zhou, Y. Li, X. He, Y. Zhang, J. Wang, X. Huang, W. Tong, J. Chen, J. Ye, C. Chen, N. Wei, G. Li, L. Dong, F. Lan, Y. Sun, Z. Zhang, Z. Yang, Y. Yu, Y. Huang, D. He, Y. Xi, D. Wei, Q. Qi, W. Li, J. Shi, M. Wang, F. Xie, J. Wang, X. Zhang, P. Wang, Y. Zhao, N. Li, N. Yang, W. Dong, S. Hu, C. Zeng, W. Zheng, B. Hao, G. sequence of red jungle fowl: Washington University School of Medicine, L. W. Hillier, S.-P. Yang, W. C. Warren, R. K. Wilson, M. evolution: Uppsala University, M. Brandstrom, H. Ellegren, B. s. a. h. W. U. genotyping, Population, R. P. M. A. Crooijmans, J. J. van der Poel, H. Bovenhuis, M. A. M. Groenen, L. Livermore National Laboratory, I. Ovcharenko, L. Gordon, L. Stubbs, D. Joint Genome Institute, S. Lucas, T. Glavina, A. Aerts, E. of application to complex traits: Institute for Animal Health, P. Kaiser, L. Rothwell, J. R. Young, S. Rogers, B. A. Walker, A. van Hateren, J. Kaufman, N. Bumstead, I. State University, S. J. Lamont, H. Zhou, R. Institute, P. M. Hocking, D. Morrice, D.-J. de Koning, A. Law, N. Bartley, D. W. Burt, U.-A. Avian Disease and Oncology Laboratory, H. Hunt, H. H. Cheng, D. and selection: Uppsala University, U. Gunnarsson, P. Wahlberg, L. Andersson, K. Institutet, E. Kindlund, M. T. Tammi, B. Andersson, H. disease genes: University of Oxford, C. Webber, C. P. Ponting, E.-b. SNP data: University of Manchester Institute of Science and Technology, I. M. Overton, P. E. Boardman, H. Tang, S. J. Hubbard, U. of Sheffield, S. A. Wilson, S. management: Beijing Institute of Genomics of Chinese Academy of Sciences, J. Yu, J. Wang et H. Yang (2004). A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* 432(7018): 717-722.
- Kao, C.-H., Z.-B. Zeng et R. D. Teasdale (1999). Multiple Interval Mapping for Quantitative Trait Loci. *Genetics* 152(3): 1203-1216.
- Kashi, Y. et D. G. King (2006). Simple sequence repeats as advantageous mutators in evolution. *Trends in Genetics* 22(5): 253-259.
- Katagiri, T., S. Asakawa, S. Minagawa, N. Shimizu, I. Hirono et T. Aoki (2001). Construction and characterization of BAC libraries for three fish species; rainbow trout, carp and tilapia. *Animal Genetics* 32(4): 200-4.

- Katagiri, T., C. Kidd, E. Tomasino, J. Davis, C. Wishon, J. Stern, K. Carleton, A. Howe et T. Kocher (2005). A BAC-based physical map of the Nile tilapia genome. *BMC Genomics* 6(1): 89.
- Kelly, P. D., F. Chu, I. G. Woods, P. Ngo-Hazelett, T. Cardozo, H. Huang, F. Kimm, L. Liao, Y.-L. Yan, Y. Zhou, S. L. Johnson, R. Abagyan, A. F. Schier, J. H. Postlethwait et W. S. Talbot (2000). Genetic Linkage Mapping of Zebrafish Genes and ESTs. *Genome Research*. 10(4): 558-567.
- Khoo, S. K., A. Ozaki, F. Nakamura, T. Arakawa, R. Nickolov, T. Sakamoto, T. Akutsu, M. Mochizuki et I. Denda (2004). Identification of a Novel Chromosomal Region Associated with Infectious Hematopoietic Necrosis (IHN) Resistance in Rainbow Trout *Oncorhynchus mykiss*. *Fish Pathology* 39(2): 95-101.
- Knapp, S. J. et W. C. Bridges (1990). Using Molecular Markers to Estimate Quantitative Trait Locus Parameters: Power and Genetic Variances for Unreplicated and Replicated Progeny. *Genetics* 126(3): 769-777.
- Knott, S. A., L. Marklund, C. S. Haley, K. Andersson, W. Davies, H. Ellegren, M. Fredholm, I. Hansson, B. Hoyheim, K. Lundstrom, M. Moller et L. Andersson (1998). Multiple Marker Mapping of Quantitative Trait Loci in a Cross Between Outbred Wild Boar and Large White Pigs. *Genetics* 149(2): 1069-1080.
- Knott, S. A. (2005). Regression-based quantitative trait loci mapping: robust, efficient and effective. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360(1459): 1435-1442.
- Korol, A. B. et K. G. Iliadi (1994). Increased recombination frequencies resulting from directional selection for geotaxis in *Drosophila*. *Heredity* 72 (Pt 1): 64-68.
- Kosambi, D. (1944). The estimation of map distance from recombination values. *Ann Eugen* 12: 172-175.
- Kota, R., S. Rudd, A. Facius, G. Kolesov, T. Thiel, H. Zhang, N. Stein, K. Mayer et A. Graner (2003). Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare*L.). *Molecular Genetics and Genomics* 270(1): 24-33.
- Kwok, P. Y., Q. Deng, H. Zakeri, S. L. Taylor et D. A. Nickerson (1996). Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* 31(1): 123-6.
- Lacape, J. M., T. B. Nguyen, S. Thibivilliers, B. Bojinov, B. Courtois, R. G. Cantrell, B. Burr et B. Hau (2003). A combined RFLP-SSR-AFLP map of tetraploid cotton based on a *Gossypium hirsutum* x *Gossypium barbadense* backcross population. *Genome* 46(4): 612-626.
- Lacoste, A., F. Jalabert, S. Malham, A. Cueff, F. Gelebart, C. Cordevant, M. Lange et S. A. Poulet (2001). A *Vibrio splendidus* strain is associated with summer mortality of juvenile oysters *Crassostrea gigas* in the Bay of Morlaix (North Brittany, France). *Disease Aquatic Organisms* 46(2): 139-45.

- Lagercrantz, U. et D. J. Lydiate (1995). RFLP Mapping in *Brassica nigra* indicates differing recombination rates in male and female meiosis. *Genome* 38: 255-264.
- Lallias, D. (2006). Genetic linkage mapping in the blue mussel *Mytilus edulis* and the European flat oyster *Ostrea edulis*, and the search for Quantitative Trait Loci of resistance to a disease in *O. edulis*. Bangor, University of Wales, College of Natural Sciences, School of Ocean Sciences.
- Lallias, D., A. R. Beaumont, C. S. Haley, P. Boudry, S. Heurtebise et S. Lapègue (2007). A first-generation genetic linkage map of the European flat oyster *Ostrea edulis* (L.) based on AFLP and microsatellite markers. *Animal Genetics* 38(6): 560-568.
- Lallias, D., S. Lapègue, C. Hecquet, P. Boudry et A. R. Beaumont (2007). AFLP-based genetic linkage maps of the blue mussel (*Mytilus edulis*). *Animal Genetics* 38(4): 340-349.
- Lander, E. S. et P. Green (1987). Construction of multilocus genetic linkage maps in Humans. *Proceedings of National Academic Sciences of the USA* 84(8): 2363-2367.
- Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly, S. E. Lincoln et L. Newburg (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1(2): 174-81.
- Lander, E. S. et D. Botstein (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121(1): 185-199.
- Lander, E. et L. Kruglyak (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* 11(3): 241-247.
- Langdon, C., F. Evans, D. Jacobson et M. Blouin (2003). Yields of cultured Pacific oysters *Crassostrea gigas* (Thunberg) improved after one generation of selection *Aquaculture* 220: 227-244.
- Launey, S. et D. Hedgecock (2001). High genetic load in the Pacific oyster *Crassostrea gigas*. *Genetics* 159(1): 255-65.
- Leal, M. S. (2003). Genetic maps of microsatellite and single-nucleotide polymorphism markers: Are the distances accurate? *Genetic Epidemiology* 24(4): 243-252.
- Lee, B.-Y., W.-J. Lee, J. T. Streebman, K. L. Carleton, A. E. Howe, G. Hulata, A. Slettan, J. E. Stern, Y. Terai et T. D. Kocher (2005). A second-generation genetic linkage map of *Tilapia (Oreochromis spp.)*. *Genetics* 170(1): 237-244.
- Lee, J. H. et X. Guo (2006). Mining EST database for single-nucleotide polymorphisms. *Journal of Shellfish Research* 25(2): 748-749.
- Leitao, A., P. Boudry, J. Labat et C. Thiriot-Quévieux (1999). Comparative karyological study of cupped oyster species. *Malacologia* 41: 175-186.

- Leitao, A., C. Thiriot-Quiévreux, P. Boudry et I. Malheiro (1999). A "G" chromosome banding study of three cupped oyster species: *Crassostrea gigas*, *Crassostrea angulata* and *Crassostrea virignica*. *Genetic Selection Evolution* 31: 519-527.
- Leitao, A., P. Boudry et C. Thiriot-Quievreux (2001). Evidence of differential chromosome loss in aneuploid karyotypes of the Pacific oyster, *Crassostrea gigas*. *Genome* 44(4): 735-737.
- Leitao, A., R. Chaves, S. Santos, H. Guedes-Pinto et P. Boudry (2004). Restriction enzyme digestion chromosome banding in *Crassostrea* and *Ostrea* species: comparative karyological analysis within Ostreidae. *Genome* 47: 781-788.
- Lenormand, T. et J. Dutheil (2005). Recombination difference between sexes: A role for haploid selection. *Plos Biology* 3(3): 396-403.
- Leroux, S. (2004). Un marché très demandeur sur le naissain d'écloserie. *L'ostréiculteur français*: 174.
- Lester, P. J. et S. J. Hubbard (2002). Comparative bioinformatic analysis of complete proteomes and protein parameters for cross-species identification in proteomics. *Proteomics* 2(10): 1392-1405.
- Letcher, B. H. et T. L. King (2001). Parentage and Grandparentage assignment with known and unknown matings: applications to Connecticut River Atlantic salmon restoration. *Canadian Journal of Fisheries and Aquatic Science* 58: 1812-1821.
- Levinson, G. et G. A. Gutman (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4(3): 203-221.
- Li, B., Q. Xia, C. Lu, Z. Zhou et Z. Xiang (2004). Analysis on frequency and density of microsatellites in coding sequences of several eukaryotic genomes. *Genomics Proteomics Bioinformatics* 2(1): 24-31.
- Li, G. et D. Hedgecock (1998). Genetic heterogeneity, detected by PCR-SSCP, among samples of larval Pacific oysters (*Crassostrea gigas*) supports the hypothesis of large variance in reproductive success. *Canadian Journal of Fisheries and Aquatic Science* 55(4): 1025-1033.
- Li, G., S. Hubert, K. Bucklin, V. Ribes et D. Hedgecock (2003). Characterization of 79 microsatellites DNA markers in the Pacific oyster *Crassostrea gigas*. *Molecular Ecology Notes* 3(2): 228-232.
- Li, L. et X. Guo (2004). AFLP-based genetic linkage maps of the pacific oyster *Crassostrea gigas* (Thunberg, 1793). *Marine Biotechnology* 6(1): 26-36.
- Li, L., J. Xiang, X. Liu, Y. Zhang, B. Dong et X. Zhang (2005). Construction of AFLP-based genetic linkage map for *Zhikong scallop*, *Chlamys farreri* and mapping of sex-linked markers. *Aquaculture* 245(1-4): 63-73.

- Li, W. H., T. Gojobori et M. Nei (1981). Pseudogenes as a paradigm of neutral evolution. *Nature* 292(5820): 237-239.
- Li, Y., K. Byrne, E. Miggiano, V. Whan, S. Moore, S. Keys, P. Crocos, N. Preston et S. Lehnert (2003). Genetic mapping of the kuruma prawn *Penaeus japonicus* using AFLP markers. *Aquaculture* 219(1-4): 143-156.
- Li, Y. C., A. B. Korol, T. Fahima, A. Beiles et E. Nevo (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology* 11(12): 2453-65.
- Liao, M., L. Zhang, G. Yang, M. Zhu, D. Wang, Q. Wei, G. Zou et D. Chen (2007). Development of silver carp (*Hypophthalmichthys molitrix*) and bighead carp (*Aristichthys nobilis*) genetic maps using microsatellite and AFLP markers and a pseudo-testcross strategy. *Animal Genetics* 38(4): 364-370.
- Lin, J. Z. et K. Ritland (1996). The effects of selective genotyping on estimates of proportion of recombination between linked quantitative trait loci. *Theoretical and Applied Genetics* 93(8): 1261-1266.
- Lincoln, S.E., A. Daly et E. S. Lander (1992). Mapping genes controlling quantitative traits with MAPMAKER/QTL. Whitehead Institute Technical Report, 2nd éd.
- Liu, B. (1998). *Genomics, Linkage Mapping and QTL Analysis*. Boca Raton, FL, CRC Press.
- Liu, Z., A. Karsi, P. Li, D. Cao et R. Dunham (2003). An AFLP-Based Genetic Linkage Map of Channel Catfish (*Ictalurus punctatus*) constructed by using an interspecific hybrid resource family. *Genetics* 165(2): 687-694.
- Livak, K. J., S. J. Flood, J. Marmaro, W. Giusti et K. Deetz (1995). Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Applications* 4(6): 357-362.
- Longwell, A. C. et S. S. Stiles (1973). Gamete cross incompatibility and inbreeding in the commercial American oyster, *Crassostrea virginica* Gmelin. *Cytologica* 38: 521-533.
- Lopez-Flores, I., R. De la Herran, M. A. Garrido-Ramos, P. Boudry, C. Ruiz-Rejon et M. Ruiz-Rejon (2004). The molecular phylogeny of oysters based on a satellite DNA related to transposons. *Gene* 15: 181-188.
- Lorch, P. D. (2005). Sex differences in recombination and mapping adaptations. *Genetica* 123(1): 39-47.
- Lorieux, M., B. Goffinet, X. Perrier, D. González de León et C. Lanaud (1995a). Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross populations. *Theoretical and Applied Genetics* 90: 73-80.
- Lorieux, M., X. Perrier, B. Goffinet, C. Lanaud et D. González de León (1995b). Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F2 populations. *Theoretical and Applied Genetics* 90: 81-89.

- Lu, B. C. (1974). Genetic Recombination in *Coprinus*. IV. a Kinetic Study of the Temperature Effect on Recombination Frequency. *Genetics* 78(2): 661-677.
- Luo, L., Y. M. Zhang et S. Xu (2005). A quantitative genetics model for viability selection. *Heredity* 94(3): 347-55.
- Lymbery, A. (2000). Genetic improvement in the Australian aquaculture industry. *Aquaculture Research* 31: 145-149.
- Lynch, M. et B. Walsh (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA.
- Lyons, E. et M. Freeling (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. *The Plant Journal* 53(4): 661-673.
- Lyttle, T. (1991). Segregation distortion. *Annual Review of Genetics* 25: 511-557.
- Mackay, T. F. (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics* 35: 303-39.
- Mackay, T. F. (2004). The genetic architecture of quantitative traits: lessons from *Drosophila*. *Current Opinion in Genetics and Development* 14(3): 253-7.
- Magoulas, A., B. Gjetvaj, V. Terzoglou et E. Zouros (1998). Three polymorphic microsatellites in the Japanese oyster *Crassostrea gigas* (Thunberg). *Animal Genetics* 29: 69-70.
- Mahon, G. A. T. (1983). Selection goals in oyster breeding. *Aquaculture* 33(1-4): 141-148.
- Manly, K. F. et J. M. Olson (1999). Overview of QTL mapping software and introduction to Map Manager QT. *Mammalian Genome* 10(4): 327-334.
- Marais, G. et L. Duret (2001). Synonymous codon usage, accuracy of translation, and gene length in *Caenorhabditis elegans*. *Journal of Molecular Evolution* 52(3): 275-280.
- Marth, G. T., I. Korf, M. D. Yandell, R. T. Yeh, Z. Gu, H. Zakeri, N. O. Stitzel, L. Hillier, P. Y. Kwok et W. R. Gish (1999). A general approach to single-nucleotide polymorphism discovery. *Nature Genetics* 23(4): 452-456.
- Martinez, O. et R. Curnow (1992). Estimating the location and the sizes of the effects of quantitative trait loci using flanking markers. *Theoretical Applied Genetics* 85: 480-488.
- Martyniuk, C. J., G. M. L. Perry, H. K. Mogahadam, M. M. Ferguson et R. G. Danzmann (2003). The genetic architecture of correlations among growth related traits and male age at maturation in rainbow trout. *Journal of Fish Biology* 63(3): 746-764.
- Matsuda, M., N. Kawato, S. Asakawa, N. Shimizu, Y. Nagahama, S. Hamaguchi, M. Sakaizumi et H. Horii (2001). Construction of a BAC library derived from the inbred Hd-rR strain of the teleost fish, *Oryzias latipes*. *Genes, Genetics and Systems* 76(1): 61-3.

- Mattiucci, S. et F. Villani (1983). Allozyme study in oysters classified as *Crassostrea gigas* (Thunberg, 1793) and *Crassostrea angulata* (Lamarck, 1819) (Mollusca: Ostreidae). *Parasitologia* 25: 21-27.
- McGoldrick, D. J., D. Hedgecock, L. English, P. Baoprasertkul et R. D. Ward (2000). The transmission of microsatellite alleles in Australian and North American stocks of the Pacific oyster (*Crassostrea gigas*): selection and null alleles. *Journal of Shellfish Research* 19: 779-788.
- McLean, K. W. et A. H. Whiteley (1973). Characteristics of DNA from the oyster, *Crassostrea gigas*. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 335: 35-41.
- Menzel, R. W. (1974). Portuguese and Japanese oysters are the same species. *Journal of the Fisheries Research Board of Canada* 31: 453-456.
- Meyers, B. C., S. V. Tingey et M. Morgante (2001). Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Research* 11(10): 1660-1676.
- Moen, T., B. Hoyheim, H. Munck et L. Gomez-Raya (2004). A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. *Animal Genetics* 35(2): 81-92.
- Moen, T., A. Sonesson, B. Hayes, S. Lien, H. Munck et T. Meuwissen (2007). Mapping of a quantitative trait locus for resistance against infectious salmon anaemia in Atlantic salmon (*Salmo Salar*): comparing survival analysis with analysis on affected/resistant data. *BMC Genetics* 8(1): 53.
- Mogahadam, H. K., J. Poissant, H. A. Fotherby, L. Haidle, M. M. Ferguson et R. G. Danzmann (2007). Quantitative trait loci for body weight, condition factor and age at sexual maturation in Arctic charr (*Salvelinus alpinus*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*). *Molecular genetics and Genomics* 277(6): 647-661.
- Mohrenweiser, H. W., S. Tsujimoto, L. Gordon et A. S. Olsen (1998). Regions of sex-specific hypo- and hyper-recombination identified through Integration of 180 Genetic Markers into the Metric Physical Map of Human Chromosome 19. *Genomics* 47(2): 153-162.
- Moran, G., J. Bell et A. Hilliker (1983). Greater meiotic recombination in male vs. female gametes in *Pinus radiata*. *Journal of Heredity* 74: 62-63.
- Morgan, T. (1914). Complete linkage in the second chromosome of the male of *Drosophila*. *Science* 36: 719-720.
- Moriyama, E. N. et J. R. Powell (1996). Intraspecific nuclear DNA variation in *Drosophila*. *Molecular Biology and Evolution* 13(1): 261-77.

- Morlais, I. et D. W. Severson (2003). Intraspecific DNA variation in nuclear genes of the mosquito *Aedes aegypti*. *Insect Molecular Biology* 12(6): 631-639.
- Morton, N. E. (1955). Sequential tests for the detection of linkage. *American Journal of Human Genetics* 7(3): 277-318.
- Naruse, K., S. Fukamachi, H. Mitani, M. Kondo, T. Matsuoka, S. Kondo, N. Hanamura, Y. Morita, K. Hasegawa, R. Nishigaki, A. Shimada, H. Wada, T. Kusakabe, N. Suzuki, M. Kinoshita, A. Kanamori, T. Terado, H. Kimura, M. Nonaka et A. Shima (2000). A Detailed Linkage Map of Medaka, *Oryzias latipes*: Comparative Genomics and Genome Evolution. *Genetics* 154(4): 1773-1784.
- Nell, J. A., A. K. Sheridan et I. R. Smith (1996). Progress in a Sydney rock oyster, *Saccostrea commercialis* breeding program. *Aquaculture* 144: 295-302.
- Nell, J. A., I. R. Smith et A. K. Sheridan (1999). Third generation evaluation of Sydney rock oyster *Saccostrea commercialis* breeding lines *Aquaculture* 170: 195-203.
- Nell, J. (2002). The Australian oyster fishery. *World Aquaculture* 33: 8-10.
- Nichols, K. M., W. P. Young, R. G. Danzmann, B. D. Robison, C. Rexroad, M. Noakes, R. B. Phillips, P. Bentzen, I. Spies, K. Knudsen, F. W. Allendorf, B. M. Cunningham, J. Brunelli, H. Zhang, S. Ristow, R. Drew, K. H. Brown, P. A. Wheeler et G. H. Thorgaard (2003). A consolidated linkage map for rainbow trout *Oncorhynchus mykiss*. *Animal Genetics* 34(2): 102-115.
- Nichols, K. M., K. W. Broman, K. Sundin, J. M. Young, P. A. Wheeler et G. H. Thorgaard (2007). Quantitative Trait Loci x maternal cytoplasmic environment interaction for development rate in *Oncorhynchus mykiss*. *Genetics* 175(1): 335-347.
- Nicolas, J. L., T. Renault, B. Gagnaire, C. Garcia, M. Garnier, M. Gay, Y. Labreuche, F. Le Roux, L. Miossec, J. F. Pépin et D. Saulnier (2007). Les risques associés aux pathogènes. Mortalités estivales de l'huître creuse *Crassostrea gigas*: Défi Morest. éd. Quae, Jean François Samain, Helen McCombie: 229-269.
- O'Foighil, D., P. M. Gaffney, A. E. Wilbur et T. J. Hilbish (1998). Mitochondrial cytochrome oxidase I gene sequences support an Asian origin for the Portuguese oyster *Crassostrea angulata*. *Marine Biology* 131: 497-503.
- O'Malley, K. G., T. Sakamoto, R. G. Danzmann et M. M. Ferguson (2003). Quantitative Trait Loci for spawning date and body weight in Rainbow Trout: testing for conserved effects across ancestrally duplicated chromosomes. *Journal of Heredity* 94(4): 273-284.
- Ozaki, A., S.-K. Khoo, Y. Yoshiura, M. Ototake, T. Sakamoto, J. M. Dijkstra et N. Okamoto (2007). Identification of additional Quantitative Trait Loci (QTL) responsible for susceptibility to infectious pancreatic necrosis virus in Rainbow Trout. *Fish Pathology* 42(3): 131-140.

- Pajot, R., F. Hennequart, P. Trintignac et P. Glize (1998). Etude de familles de naissains pour la compréhension de mortalités anormales chez l'huître creuse *Crassostrea gigas* Rapport SMIDAP. Nantes, France: 36p.
- Palti, Y., J. E. Parsons et G. H. Thorgaard (1999). Identification of candidate DNA markers associated with IHN virus resistance in backcrosses of rainbow (*Oncorhynchus mykiss*) and cutthroat trout (*O. clarki*). *Aquaculture* 173(1-4): 81-94.
- Palti, Y., S. A. Gahr, J. D. Hansen et C. E. Rexroad, 3rd (2004). Characterization of a new BAC library for rainbow trout: evidence for multi-locus duplication. *Animal Genetic* 35(2): 130-3.
- Papachristou, C. et S. Lin (2006). Microsatellites versus Single-Nucleotide Polymorphisms in confidence interval estimation of disease loci. *Genetic Epidemiology* 30(1): 3-17.
- Pardo-Manuel de Villena, F., E. de la Casa-Esperon, T. L. Briscoe et C. Sapienza (2000). A genetic test to determine the origin of maternal transmission ratio distortion: meiotic drive at the mouse Om locus. *Genetics* 154(1): 333-342.
- Pardo-Manuel de Villena, F. et C. Sapienza (2001). Nonrandom segregation during meiosis: the unfairness of females. *Mammalian Genome* 12(5): 331-339.
- Paterson, A., S. Damon, J. Hewitt, D. Zamir, R. HD, S. E. Lincoln, E. Lander et S. D. Tanksley (1991). Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. *Genetics* 127: 181-187.
- Peden, J. F. (1999). CodonW: Analysis of Codon Usage. University of Nottingham.
- Pépin, J.-F., A. Riou et T. Renault (2008). Rapid and sensitive detection of ostreid herpesvirus 1 in oyster samples by realtime PCR. *Journal of Virological Methods* 149(2): 269-276.
- Perez, F., C. Erazo, M. Zhinaula, F. Volckaert et J. Calderon (2004). A sex-specific linkage map of the white shrimp (*Penaeus Litopenaeus vannamei*) based on AFLP markers. *Aquaculture* 242(1-4): 105-118.
- Perry, G. M. L., R. G. Danzmann, M. M. Ferguson et J. P. Gibson (2001). Quantitative trait loci for upper thermal tolerance in outbred strains of rainbow trout (*Oncorhynchus mykiss*). *Heredity* 83: 333-341.
- Peyre, J. F. L., R. K. Copper, J. E. Supan et A. K. Volety (1999). Total bacteria and *Vibrio vulnificus* load in diploid and triploid eastern oysters in Louisiana. *Journal of Shellfish Research* 18(1): 324-325.
- Philipp, U., P. Wehling et G. Wricke (1994). A linkage map of rye. *Theoretical Applied Genetics* 88: 243-248.
- Phillips, B. T., K. Bolding et B. B. Riley (2001). Zebrafish fgf3 and fgf8 encode redundant functions required for otic placode induction. *Development Biology* 235(2): 351-65.

- Phillips, R. B., A. Zimmerman, M. A. Noakes, Y. Palti, M. R. Morasch, L. Eiben, S. S. Ristow, G. H. Thorgaard et J. D. Hansen (2003). Physical and genetic mapping of the rainbow trout major histocompatibility regions: evidence for duplication of the class I region. *Immunogenetics* 55(8): 561-9.
- Phillips, R. B., M. R. Morasch, L. K. Park, K. A. Naish et R. H. Devlin (2005). Identification of the sex chromosome pair in coho salmon (*Oncorhynchus kisutch*): lack of conservation of the sex linkage group with chinook salmon (*Oncorhynchus tshawytscha*). *Cytogenetic Genome Research* 111(2): 166-70.
- Picoult-Newberg, L., T. E. Ideker, M. G. Pohl, S. L. Taylor, M. A. Donaldson, D. A. Nickerson et M. Boyce-Jacino (1999). Mining SNPs from EST databases. *Genome Research* 9(2): 167-74.
- Plomin, R., G. E. McClearn, D. L. Smith, S. Vignetti, M. J. Chorney, K. Chorney, C. P. Venditti, S. Kasarda, L. A. Thompson, D. K. Detterman et al. (1994). DNA markers associated with high versus low IQ: the IQ Quantitative Trait Loci (QTL) Project. *Behavioural Genetics* 24(2): 107-18.
- Plomion, C. (2005). Les marqueurs génétiques moléculaires: Méthodes de détection et Applications. <http://www.pierroton.inra.fr/genetics/cartoqtl/>
- Primmer, C. R., N. Saino, A. P. Moller et H. Ellegren (1996). Directional evolution in germline microsatellite mutations. *Nature Genetic* 13(4): 391-393.
- Qin, Y., X. Liu, H. Zhang, G. Zhang et X. Guo (2007). Genetic mapping of size-related quantitative trait loci (QTL) in the bay scallop (*Argopecten irradians*) using AFLP and microsatellite markers. *Aquaculture* 272(1-4): 281-290.
- Qin, Y., X. Liu, H. Zhang, G. Zhang et X. Guo (2007). Identification and mapping of amplified fragment length polymorphism markers linked to shell color in Bay Scallop, *Argopecten irradians* (Lamarck, 1819). *Marine Biotechnology* 9(1): 66-73.
- Quilang, J., S. Wang, P. Li, J. Abernathy, E. Peatman, Y. Wang, L. Wang, Y. Shi, R. Wallace, X. Guo et Z. Liu (2007). Generation and analysis of ESTs from the eastern oyster, *Crassostrea virginica* Gmelin and identification of microsatellite and SNP markers. *BMC Genomics* 8: 157.
- Quiniou, S. M., T. Katagiri, N. W. Miller, M. Wilson, W. R. Wolters et G. C. Waldbieser (2003). Construction and characterization of a BAC library from a gynogenetic channel catfish *Ictalurus punctatus*. *Genetic Selection Evolution* 35(6): 673-83.
- Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology* 5(2): 94-100.
- Ranson, G. (1948). Prodissoconques et classification des ostréides vivants. *Bulletin du Musée Royal d'Histoire Naturelle de Belgique* 24: 1-12.
- Ranson, G. (1960). Les prodissoconques (coquilles larvaires) des ostréides vivants. *Bulletin de l'Institut Océanographique de Monaco* 1: 1-41.

- Rebai, A., B. Goffinet et B. Mangin (1994). Approximate thresholds of interval mapping tests for QTL detection. *Genetics* 138(1): 235-40.
- Reece, K. S., W. L. Ribeiro, P. M. Gaffney, R. B. Carnegie et S. K. Allen, Jr. (2004). Microsatellite marker development and analysis in the eastern oyster (*Crassostrea virginica*): confirmation of null alleles and non-Mendelian segregation ratios. *Journal of Heredity* 95(4): 346-352.
- Reid, D. P., A. Szanto, B. Glebe, R. G. Danzmann et M. M. Ferguson (2005). QTL for body weight and condition factor in Atlantic salmon (*Salmo salar*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). *Heredity* 94(2): 166-172.
- Reiter, R. S., J. Williams, K. Feldman, A. J. Rafalski, S. V. Tingey et P. A. Scolnik (1992). Global and local genome mapping in *Arabidopsis thaliana* recombinant inbred lines and random amplified polymorphic DNAs. *Proceedings of National Academic Science of the USA* 89: 1477-1481.
- Renault, T., R. M. Le Deuff, N. Cochenec, B. Chollet et P. Maffart (1995). Herpes-like viruses associated with high mortality levels in larvae and spat of Pacific oysters, *Crassostrea gigas*: a comparative study, the thermal effects on virus detection in hatchery-reared larvae, reproduction of the disease in axenic larvae. *Veterinary Research* 26(5-6): 539-43.
- Renault, T., C. Lipart et I. Arzul (2001). A herpes-like virus infects a non-ostreid bivalve species: virus replication in *Ruditapes philippinarum* larvae. *Diseases of Aquatic Organisms* 45(1): 1-7.
- Rick, C. (1969). Controlled introgression of chromosomes of *Solanum pennellii* into *Lycopersicon esculentum*: segregation and recombination. *Genetics* 62: 753-768.
- Robert, R. et A. Gérard (1999). Bivalve hatchery technology: the current situation for the pacific oyster *Crassostrea gigas* and the scallop *Pecten maximus* in France. *Aquatic Living Resources* 12: 121-130.
- Robertson, D. (1984). A possible technique for isolating genic DNA for quantitative traits in plants. *Journal of Theoretical Biology* 117: 1-10.
- Robison, B. D., P. A. Wheeler, K. Sundin, P. Sikka et G. H. Thorgaard (2001). Composite interval mapping reveals a major locus influencing embryonic development rate in rainbow trout (*Oncorhynchus mykiss*). *Journal of Heredity* 92(1): 16-22.
- Robledo, J. A. F., W. T. Pecher et G. R. Vasta (2000). High-throughput isolation of oyster DNA facilitates diagnosis of "Dermo" disease. *Qiagen News* 5: 15-17.
- Rodriguez, M. F., S. LaPatra, S. Williams, T. Famula et B. May (2004). Genetic markers associated with resistance to infectious hematopoietic necrosis in rainbow and steelhead trout (*Oncorhynchus mykiss*) backcrosses. *Aquaculture* 241(1-4): 93-115.

- Rohrer, G. A., L. J. Alexander, Z. Hu, T. P. Smith, J. W. Keele et C. W. Beattie (1996). A comprehensive map of the porcine genome. *Genome Research* 6(5): 371-391.
- Ronin, Y. I., A. B. Korol et J. I. Weller (1998). Selective genotyping to detect quantitative trait loci affecting multiple traits: interval mapping analysis. *Theoretical and Applied Genetics* 97(7): 1169-1178.
- Rose, A. et D. Baillie (1979). The effect of temperature and parental age on recombination and non disjunction in *Caenorhabditis elegans*. *Genetics* 92: 409-412.
- Rozas, J. et R. Rozas (1995). DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Computational Applied Biosciences* 11(6): 621-625.
- Rozen, S. et H. Skaletsky (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Molecular Biology* 132: 365-386.
- Sachidanandam, R., D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein, G. Marth, S. Sherry, J. C. Mullikin, B. J. Mortimore, D. L. Willey, S. E. Hunt, C. G. Cole, P. C. Coggill, C. M. Rice, Z. Ning, J. Rogers, D. R. Bentley, P. Y. Kwok, E. R. Mardis, R. T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R. H. Waterston, J. D. McPherson, B. Gilman, S. Schaffner, W. J. Van Etten, D. Reich, J. Higgins, M. J. Daly, B. Blumenstiel, J. Baldwin, N. Stange-Thomann, M. C. Zody, L. Linton, E. S. Lander et D. Altshuler (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409 (6822): 928-33.
- Sakamoto, T., R. G. Danzmann, N. Okamoto, M. M. Ferguson et P. E. Ihssen (1999). Linkage analysis of quantitative trait loci associated with spawning time in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 173(1-4): 33-43.
- Sakamoto, T., R. G. Danzmann, K. Gharbi, P. Howard, A. Ozaki, S. K. Khoo, R. A. Woram, N. Okamoto, M. M. Ferguson, L.-E. Holm, R. Guyomard et B. Hoyheim (2000). A microsatellite linkage map of Rainbow Trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* 155(3): 1331-1345.
- Samain, J. F. et H. McCombie (2007). Mortalités estivales de l'huître creuse *Crassostrea gigas*. *Défi Morest, Ifremer/Quae*.
- SanCristobal, M., G. Renand, Y. Amigues, M.-Y. Boshier, H. Levéziel et B. Bibé (2000). Traçabilité individuelle des viandes bovines à l'aide de marqueurs génétiques. *INRA Production Animales* 13: 239-246.
- Sapienza, C. et W. F. Doolittle (1982). Unusual physical organization of the Halobacterium genome. *Nature* 295(5848): 384-9.
- Sauvage, C., N. Bierne, S. Lapègue et P. Boudry (2007). Single nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. *Gene* 406(1-2): 13-22.

- Sauvage, C., P. Boudry et S. Lapègue (2008). Identification and characterization of 18 novel polymorphic microsatellite makers derived from expressed sequence tags in the Pacific oyster *Crassostrea gigas*. *Molecular Ecology Notes*. Accepted.
- Sax, K. (1923). The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* 8: 552-560.
- Schiex, T. et C. Gaspin (1997). CARTHAGENE: constructing and joining maximum likelihood genetic maps. *Proceedings of International Conference on Intelligent Systems for Molecular Biology* 5: 258-67.
- Schlötterer, C. et D. B. Goldstein (1999). *Microsatellites : Evolution and Applications*. Oxford, Oxford University Press.
- Schneeberger, C., P. Speiser, F. Kury et R. Zeillinger (1995). Quantitative detection of reverse transcriptase-PCR products by means of a novel and sensitive DNA stain. *PCR Methods Applied* 4(4): 234-238.
- Schon, C. C., H. F. Utz, S. Groh, B. Truberg, S. Openshaw et A. E. Melchinger (2004). Quantitative Trait Locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* 167(1): 485-498.
- Schwarz-Sommer, Z., E. de Andrade Silva, R. Berndtgen, W.-E. Lonngig, A. Muller, I. Nindl, K. Stuber, J. Wunder, H. Saedler, T. Gubitza, A. Borking, J. F. Golz, E. Ritter et A. Hudson (2003). A Linkage Map of an F2 Hybrid Population of *Antirrhinum majus* and *A. molle*. *Genetics* 163(2): 699-710.
- Seaton, G., C. S. Haley, S. A. Knott, M. Kearsey et P. M. Visscher (2002). QTL Express: mapping quantitative trait loci in simple and complex pedigrees. *Bioinformatics* 18: 339-340.
- Sekino, M., M. Hamaguchi, F. Aranishi et K. Okoshi (2003). Development of novel microsatellite DNA markers from the Pacific Oyster *Crassostrea gigas*. *Marine Biotechnology* 5(3): 227-233.
- Sekino, M. et M. Hara (2007). Linkage maps for the Pacific Abalone (*Genus Haliotis*) based on microsatellite DNA markers. *Genetics* 175(2): 945-958.
- Sellos, D., J. Moal, L. Degremont, A. Huvet, J. Y. Daniel, S. Nicoulaud, P. Boudry, J. F. Samain et A. Van Wormhoudt (2003). Structure of amylase genes in populations of Pacific Cupped oyster (*Crassostrea gigas*): tissue expression and allelic polymorphism. *Mar Biotechnol (NY)* 5(4): 360-372.
- Sherry, S. T., M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski et K. Sirotkin (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research* 29(1): 308-11.

- Shields, D. C., A. Collins, K. H. Buetow et N. E. Morton (1991). Error filtration, interference, and the human linkage map. *Proceedings of National Academic Sciences of the USA* 88(15): 6501-6505.
- Sillanpaa, M. J. et E. Arjas (1998). Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics* 148: 1373-1388.
- Singer, A., H. Perlman, Y. Yan, C. Walker, G. Corley-Smith, B. Brandhorst et J. Postlethwait (2002). Sex-specific recombination rates in zebrafish (*Danio rerio*). *Genetics* 160(2): 649-57.
- Singh, S. et E. Zouros (1978). Genetic variation associated with growth rate in the American oyster (*Crassostrea virginica*). *Evolution* 32: 342-353.
- Small, K. S., M. Brudno, M. M. Hill et A. Sidow (2007). Extreme genomic variation in a natural population. *Proceedings of the National Academy of Sciences* 104(13): 5698-5703.
- Snelling, W., M. Gautier, J. Keele, T. Smith, R. Stone, G. Harhay, G. Bennett, N. Ihara, A. Takasuga, H. Takeda, Y. Sugimoto et A. Eggen (2004). Integrating linkage and radiation hybrid mapping data for bovine chromosome 15. *BMC Genomics* 5(1): 77.
- Soller, M., A. Genizi et T. Brody (1976). On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theoretical Applied Genetics* 47: 35-39.
- Song, K., M. Slocum et T. Osborn (1995). Molecular markers analysis of genes controlling morphological variation in *Brassica rapa*. *Theoretical Applied Genetic* 90: 1-10.
- Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *The Plant Journal* 3: 739-744.
- Sun, X. et L. Liang (2004). A genetic linkage map of common carp (*Cyprinus carpio L.*) and mapping of a locus associated with cold tolerance. *Aquaculture* 238(1-4): 165-172.
- Tadmor, Y., D. Zamir et G. Ladizinsky (1987). Genetic mapping of an ancient translocation in the genus *Lens*. *Theoretical Applied Genetic* 73: 883-892.
- Tanguy, A., I. Boutet, F. Bonhomme, P. Boudry et D. Moraga (2002). Polymorphism of metallothionein genes in the Pacific oyster *Crassostrea gigas* as a biomarker of response to metal exposure. *Biomarkers* 7: 439-450.
- Tanguy, A., X. Guo et S. E. Ford (2004). Discovery of genes expressed in response to *Perkinsus marinus* challenge in Eastern (*Crassostrea virginica*) and Pacific (*C. gigas*) oysters. *Gene* 338(1): 121-131.
- Tanguy, A., I. Boutet, J. Laroche et D. Moraga (2005). Molecular identification and expression study of differentially regulated genes in the Pacific oyster *Crassostrea gigas* in response to pesticide exposure. *FEBS Journal* 272(2): 390-403.

- Tanguy, A., N. Bierne, C. Saavedra, B. Pina, E. Bachere, M. Kube, E. Bazin, F. Bonhomme, P. Boudry, V. Boulo, I. Boutet, L. Cancela, C. Dossat, P. Favrel, A. Huvet, S. Jarque, D. Jollivet, S. Klages, S. Lapegue, R. Leite, J. Moal, D. Moraga, R. Reinhardt, J.-F. Samain, E. Zouros et A. Canario (2008). Increasing genomic information in bivalves through new EST collections in four species: Development of new genetic markers for environmental studies and genome evolution. *Gene* 408(1-2): 27-36.
- Tanksley, S. D., M. W. Ganal, J. P. Prince, M. C. de Vicente, M. W. Bonierbale, P. Broun, T. M. Fulton, J. J. Giovannoni, S. Grandillo, G. B. Martin et al. (1992). High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132(4): 1141-1160.
- Taris, N. (2005). Conséquences génétiques de la production intensive de larves d'huîtres en éclosérie: Etude des processus de dérive et de sélection liés aux pratiques d'élevage. Thèse de doctorat. Université de La Rochelle: 205 pp.
- Tenaillon, M. I., M. C. Sawkins, A. D. Long, R. L. Gaut, J. F. Doebley et B. S. Gaut (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays ssp. mays L.*). *Proceedings of National Academic Science U S A* 98(16): 9161-9166.
- Thiriot-Quévieux, C., J. Soyer, F. de Bovée et P. Albert (1988). Unusual chromosome complement in the brooding bivalve *Lasaea consanguinea*. *Genetica* 76: 143-151.
- Thiriot-Quévieux, C., G. H. Pogson et E. Zouros (1992). Genetics of growth rate variation in bivalves: aneuploidy and heterozygosity effects in a *Crassostrea* family. *Genome* 35: 39-45.
- Thoday, J. (1961). Location of polygenes. *Nature* 191: 368-370.
- Thornsberry, J. M., M. M. Goodman, J. Doebley, S. Kresovich, D. Nielsen et E. S. t. Buckler (2001). Dwarf8 polymorphisms associate with variation in flowering time. *Nature Genetic* 28(3): 286-289.
- Thorsen, J., B. Zhu, E. Frengen, K. Osoegawa, P. J. de Jong, B. F. Koop, W. S. Davidson et B. Hoyheim (2005). A highly redundant BAC library of Atlantic salmon (*Salmo salar*): an important tool for salmon projects. *BMC Genomics* 6(1): 50.
- Till, B. J., C. Burtner, L. Comai et S. Henikoff (2004). Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Res* 32(8): 2632-2641.
- Tinker, N.A. et D.E. Mather (1995). MQTL: software for simplified composite interval mapping of QTL in multiple environments. *Journal of Quantitative Trait Loci* 1.
- Toro, J. E. et G. F. Newkirk (1990). Divergent selection for growth rate in the European oyster *Ostre edulis*: response to selection and estimation of genetic parameters. *Marine Ecology Progress Series*. 62: 219-227.
- Tost, J. et I. G. Gut (2002). Genotyping single nucleotide polymorphisms by mass spectrometry. *Mass Spectrom Review* 21(6): 388-418.

- Toth, G., Z. Gaspari et J. Jurka (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research* 10(7): 967-981.
- Towe, A. M. et D. R. Stadler (1964). Effects of temperature on crossing over in *Neurospora*. *Genetics* 49(4): 577-583.
- Tulsieram, L., J. Glaubitz, G. Kiss et J. Carlson (1992a). Single tree genetic linkage analysis in conifers using haploid DNA from megagametophytes. *Biotechnology* 10: 686-690.
- Utz, H.F. et A.E. Melchinger (1996). PLABQTL: a program for composite interval mapping of QTL. *Journal of Quantitative Trait Loci* 2.
- Vaiman, D. (2000). Etablissement des cartes génétiques. INRA Production Animales Hors Série "Génétique Moléculaire": 73-78.
- Van Haeringen, W. A., M. Den Bieman, G. F. Gillissen, A. Lankhorst, M. T. R. Kuiper, L. F. M. Van Zutphen et H. A. Van Lith (2001). Mapping of a QTL for serum HDL cholesterol in the rabbit using AFLP technology. *Journal of Heredity* 92(4): 322-326.
- Van Laere, A. S., M. Nguyen, M. Braunschweig, C. Nezer, C. Collette, L. Moreau, A. L. Archibald, C. S. Haley, N. Buys, M. Tally, G. Andersson, M. Georges et L. Andersson (2003). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* 425(6960): 832-6.
- Van Ooijen, J.W. (2004). MapQTL 5, Software for the mapping of quantitative trait loci in experimental populations. Kyazma, B. V., Wageningen, Netherlands.
- Vignal, A., D. Milan, M. SanCristobal et A. Eggen (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* 34: 275-305.
- Visscher, P. M., R. Thompson et C. S. Haley (1996). Confidence intervals in QTL mapping by bootstrapping. *Genetics* 143(2): 1013-20.
- Vizir, I. et A. B. Korol (1990). Sex difference in recombination frequency in *Arabidopsis*. *Heredity* 65: 379-383.
- Volckaert, F. A. M. J., C. Batargias, D. Chatziplis, D. Chistiakov, Y. Guo, C. S. Haley, B. Hellemans, G. Kotoulas, A. Law, A. Libertini, B. J. McAndrew, C. Tsigenopoulos et H. A. Whitaker (2007). A second generation linkage map of European sea bass. *Aquaculture* 272(Supplement 1): S316.
- Voorrips, R. E. (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity*(1): 77-78.
- Wada, H., K. Naruse, A. Shimada et A. Shima (1995). Genetic linkage map of a fish, the Japanese medaka *Oryzias latipes*. *Molecular Biology and Biotechnology* 4(3): 269-274.

- Waldbieser, G., B. G. Bosworth, D. J. Nonneman et W. R. Wolters (2001). A Microsatellite-Based Genetic Linkage Map for Channel Catfish, *Ictalurus punctatus*. *Genetics* 158(2): 727-734.
- Walling, G. A., P. M. Visscher, L. Andersson, M. F. Rothschild, L. Wang, G. Moser, M. A. M. Groenen, J.-P. Bidanel, S. Cepica, A. L. Archibald, H. Geldermann, D. J. de Koning, D. Milan et C. S. Haley (2000). Combined analyses of data from Quantitative Trait Loci mapping studies: chromosome 4 effects on porcine growth and fatness. *Genetics* 155(3): 1369-1378.
- Walne, P. R. 1974. Culture of bivalve molluscs: 50 years' experience at Conwy. Fishing News (Books) Ltd., Surrey, England. 173 p.
- Wang, Y., Z. Xu et X. Guo (2001). A centromeric satellite sequence in the Pacific oyster (*Crassostrea gigas*, Thunberg) identified by fluorescence in situ hybridization. *Marine Biotechnology* (NY) 3(5): 486-92.
- Wang, L., L. Song, Y. Chang, W. Xu, D. Ni et X. Guo (2005). A preliminary genetic map of Zhikong scallop (*Chlamys farreri*, Jones and Preston 1904). *Aquaculture Research* 36(7): 643-653.
- Wang, Y., R. Ren et Z. Yu (2008). Bioinformatic mining of EST-SSR loci in the Pacific oyster, *Crassostrea gigas*. *Animal Genetics*, On line.
- Whitkus, R. (1998). Genetics of adaptive radiation in Hawaiian and Cook Islands species of *Tetramolopium* (Asteraceae). II. Genetic linkage map and its implications for interspecific breeding barriers. *Genetics* 150(3): 1209-16.
- Whitaker, H. A., B. J. McAndrew et J. B. Taggart (2006). Construction and characterization of a BAC library for the European sea bass *Dicentrarchus labrax*. *Animal Genetics* 37(5): 526-526.
- Wilson, K., Y. Li, V. Whan, S. Lehnert, K. Byrne, S. Moore, S. Pongsomboon, A. Tassanakajon, G. Rosenberg, E. Ballment, Z. Fayazi, J. Swan, M. Kenway et J. Benzie (2002). Genetic mapping of the black tiger shrimp *Penaeus monodon* with amplified fragment length polymorphism. *Aquaculture* 204(3-4): 297-309.
- Wintero, A. K., M. Fredholm et P. D. Thomsen (1992). Variable (dG-dT)_n.(dC-dA)_n sequences in the porcine genome. *Genomics* 12(2): 281-288.
- Wolff, R. K., R. Plaetke, A. J. Jeffreys et R. White (1989). Unequal crossingover between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. *Genomics* 5(2): 382-384.
- Xu, Y. et J. H. Crouch (2008). Marker-assisted selection in plant breeding: from publications to practice. *Crop Science Society of America* 48: 391-407.
- Yamtich, J., M. L. Voigt, G. Li et D. Hedgecock (2005). Eight microsatellite loci for the Pacific oyster *Crassostrea gigas*. *Animal Genetics* 36(6): 524-526.

- Young, W. P., P. A. Wheeler, V. H. Coryell, P. Keim et G. H. Thorgaard (1998). A detailed linkage map of Rainbow Trout produced using doubled haploids. *Genetics* 148(2): 839-850.
- Yu, H. et Q. I. Li (2007). EST-SSR markers from the Pacific oyster, *Crassostrea gigas*. *Molecular Ecology Notes* 7(5): 860-862.
- Yu, H. et Q. Li (2008). Exploiting EST databases for the development and characterization of EST-SSRs in the Pacific Oyster (*Crassostrea gigas*). *Journal of Heredity* 99: 208-214.
- Yu, Z. et X. Guo (2003). Genetic linkage map of the Eastern Oyster *Crassostrea virginica* Gmelin. *Biology Bulletin* 204(3): 327-338.
- Yu, Z. et X. Guo (2006). Identification and mapping of disease-resistance QTLs in the eastern oyster, *Crassostrea virginica* Gmelin. *Aquaculture* 254: 160-170.
- Zahn, L. M. et D. J. Kwiatkowski (1995). A 37-Marker PCR-based genetic linkage map of Human chromosome 9: observations on mutations and positive interference. *Genomics* 28(2): 140-146.
- Zane, L., L. Bargelloni et T. Patarnello (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology* 11(1): 1-16.
- Zeng, Z. B. (1993). Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proceedings of National Academic Sciences of the USA* 90(23): 10972-6.
- Zeng, Z. B. (1994). Precision mapping of Quantitative Trait Loci. *Genetics* 136(4): 1457-1468.
- Zhang, L., C. Yang, Y. Zhang, L. Li, X. Zhang, Q. Zhang et J. Xiang (2007). A genetic linkage map of Pacific white shrimp (*Litopenaeus vannamei*): sex-linked microsatellite markers and high recombination rates. *Genetica* 131(1): 37-49.
- Zhu, C., C. Wang et Y.-M. Zhang (2007). Modeling segregation distortion for viability selection I. Reconstruction of linkage maps with distorted markers. *Theoretical and Applied Genetics* 114(2): 295-305.
- Zhu, Y., J. E. Strassmann et D. C. Queller (2000). Insertions, substitutions, and the origin of microsatellites. *Genetic Research* 76(3): 227-236.
- Zhuchenko, A. et A. B. Korol (1983). Ecological aspect of recombination problem. *Theoretical Applied Genetics* 64: 117-185.
- Zimmerman, A. M., P. A. Wheeler, S. S. Ristow et G. H. Thorgaard (2005). Composite interval mapping reveals three QTL associated with pyloric caeca number in rainbow trout, *Oncorhynchus mykiss*. *Aquaculture* 247: 85-95.
- Zouros, E. et G. H. Pogson (1994). Heterozygosity, heterosis and adaptation. *Genetics and Evolution of Aquatic Organisms*. London, A. R. Beaumont. Chapman & Hall, 135-146.

Zouros, E., C. Thiriot-Quiévreux et G. Kotoulas (1996). The negative correlation between somatic aneuploidy and growth in the oyster *Crassostrea gigas* and implications for the effects of induced polyploidization. *Genetic Research* 68: 109-116.

LISTE DES FIGURES

ET DES TABLEAUX

Figures :

Figure 1 : Cycle biologique de l'huître creuse *Crassostrea gigas*

Figure 2 : Anatomie à l'âge adulte de l'huître creuse *Crassostrea gigas*

Figure 3 : Principaux pays producteur de l'huître creuse *C.gigas*

Figure 4 : Production annuelle mondiale de l'huître creuse (en tonnes)

Figure 5 : Répartition naturelle mondiale de l'huître creuse *C. gigas* et zones d'introduction

Figure 6 : Répartition de la production de l'huître creuse sur les sites d'élevage français

Figure 7 : Taille de génomes chez les vertébrés et invertébrés (pg)

Figure 8 : Estimation de la taille de génomes chez quelques bivalves (pg)

Figure 9 : Pourcentages moyens de mortalité et écart-types des 17 familles de demi-frères relevés dans les 3 sites le 02 octobre 2001.

Figure 10 : Moyennes et écarts types par mâle de la génération «0» des mortalités des lots G₃cR et G₃cS en octobre 2003 sur le site de Ronce

Figure 11 : Corrélation entre les mortalités cumulées au cours de l'expérience L4 et les mortalités à Ronce sur estran des lots G₃c².

Figure 12 : Plan de croisement à l'origine des individus F0 dans le cadre du projet Aquafirst.

Figure 13 : Schéma synthétique de la production des descendants F2 et du suivi de mortalité en 2006.

Figure 14 : Profil thermique appliqué au cours de la réaction de PCR-Q

Figure 15 : Utilisation de la courbe standard pour déduire la quantité d'ADN viral présente dans l'échantillon

Figure 16 : Cinétiques de mortalité journalière des cinq familles F2 étudiées au cours de la période estivale 2006 (n=1000/famille).

Figure 17 : pourcentages cumulés de mortalité des familles F2 étudiées au cours de la période estivale 2006 (Juillet à octobre 2006, n=1000/famille).

Figure 18 : Représentation graphique en Box-plot de la variance de la charge virale au sein des trois familles F2 et entre conditions

Figure 19 : Schématisation du principe de détection du polymorphisme de microsatellites en ségrégation dans des familles.

Figure 20 : Polymorphisme de séquence : (a) Polymorphisme de séquence d'un seul nucléotide (SNP) ; (b) Polymorphisme d'insertion ou de délétion (Indel).

Figure 21 : Principales techniques de génotypage SNP (d'après Plomion, 2005)

Figure 22 : Principe de la discrimination allélique par hybridation allèle spécifique de sondes TaqMan. Le fluorophore donneur est représenté en jaune et le fluorophore receveur en noir

Figure 23 : Description des étapes du processus de génotypage par technologie SNPlex

Figure 24 : Cas idéal de marqueur SNP pleinement informatif (pas d'ambiguïté sur l'origine parentale de l'allèle C).

Figure 25 : Répartition et nature des marqueurs SNP identifiés

Figure 26 : Pourcentage de transitions et de transversions dans les régions codantes et non codantes

Figure 27 : Différents types de cartes (liaison, physique, d'hybride de radiation et cytogénétique)

Figure 28 : Fonctions de cartographies linéaires

Figure 29: Schématisation du phénomène de biais méiotique

Figure 30 : Evolution du nombre de publications de cartes génétiques chez les espèces d'intérêt aquacole au cours de la dernière décennie.

Figure 31 : Cartographie génétique consensus composée de 10 groupes de liaison, comportant 80 marqueurs et d'une longueur totale de 1062cM

Figure 32 : Corrélacion entre le nombre de marqueurs et la Taille du groupe de liaison (cM)

Figure 33 : Cartes génétiques consensus mâle et femelle

Figure 34 : Corrélacion de la distance inter-marqueur entre paires de marqueurs communes aux carte consensus et carte publiée par Hubert et Hedgecock (2004)

Figure 35 : Primo-détection de QTL dans les groupes de liaison II, V, VI, VII et IX

Figure 36 : Localisation des cinq QTL dans les groupes de liaison V, VI, VII et IX

Figure 37 : Mise en évidence d'un faux QTL au sein du groupe de liaison II

Figure 38 : Part de variance phénotypique lié au phénotype « survie ».

Figure 39 : Composante phénotypique maternelle (Dam effect - Module Half-Sib)

Figure 40 : Composante phénotypique paternel (Sire effect - Module Half-Sib)

Figure 41 : Ségrégation différentielle des QTL dans les familles F2 (les QTL significatifs sont encadrés)

Figure 42 : Schéma conventionnel de validation des QTN depuis les QTL

Tableaux :

Tableau 1 : Principaux EST développés chez *C. gigas*.

Tableau 2 : Principales caractéristiques des deux cartes de liaison développées chez *C.gigas*.

Tableau 3: Estimations d'héritabilité réalisée pour la survie du naissain estimée en troisième génération en octobre 2003

Tableau 4 : Ensemble des suivis réalisés pour l'étude de la survie en laboratoire

Tableau 5 : Plan de croisements F2

Tableau 6 : Composition du Mix de PCR quantitative pour la quantification du virus OsHV1

Tableau 7 : Estimation du taux d'éclosion à J+2 et de pourcentage de larves fixées.

Tableau 8 : Echantillons prélevés puis congelés au cours du suivi de mortalité

Tableau 9 : Statistiques descriptives de la charge virale par famille F2 et par condition (Individus morts, vivants ou survivants)

Tableau 10 : Caractéristiques des principaux marqueurs moléculaires

Tableau 11 : Caractéristiques des marqueurs microsatellites génotypés

Tableau 12 : SNP génotypés dans chaque famille F2

Tableau 13 : Cartographies publiées chez les espèces d'intérêt aquacole

Tableau 14 : Longueur, intervalle minimum, moyen et maximum et constitution des groupes de liaison.

Tableau 15 : Comparatif du nombre de marqueur, de la longueur et de l'intervalle pour chaque groupe de liaison

Tableau 16 : Principaux logiciels de détection de QTL

Tableau 17 : Revue des QTL détectés chez les principales espèces d'intérêt aquacole

Tableau 18 : Part de variance phénotypique respective expliquant la variation des deux caractères.

Tableau 19 : Lod score, positions estimées et intervalles de confiance à 95% des régions QTL pour chaque caractère.

ANNEXES

Annexe 1: Statistiques de la recherche de QTL parmi les 10 groupes de liaison

LG II					
Mortalité			Charge Virale		
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
11.07	21.49	4.666	14.22	27.39	5.947
Source	Estimation	Ecart type.	Source	Estimation	Ecart type
Moyenne	0.4817	0.042785	Moyenne	3.5741	0.211662
Effet additif	0.1945	0.041439	Effet additif	1.0911	0.205003
Effet de Dominance	0.0316	0.06751	Effet de Dominance	0.1693	0.333979

LG V					
Mortalité			Charge Virale		
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
8.79	17.4	3.779	9.89	19.56	4.248
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.4163	0.042532	Moyenne	3.0635	0.216216
Effet additif	0.1407	0.033996	Effet additif	0.7651	0.172824
Effet de Dominance	0.1324	0.078785	Effet de Dominance	0.6038	0.40051

LG VI					
Mortalité			Charge Virale		
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
13.58	26.76	5.81	14.91	29.33	6.368
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.591	0.034575	Moyenne	4.0876	0.174549
Effet additif	0.0879	0.037046	Effet additif	0.2611	0.187022
Effet de Dominance	-0.189	0.055463	Effet de Dominance	-1.223	0.279999

LG VII					
Mortalité			Charge Virale		
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
10.94	21.6	4.691	5.26	10.46	2.272
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.6414	0.035478	Moyenne	3.9769	0.181559
Effet additif	-0.0263	0.039696	Effet additif	-0.1164	0.203142
Effet de Dominance	-0.2533	0.055237	Effet de Dominance	-0.8924	0.282672

LG IX					
Mortalité			Charge Virale		
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
8.6	17.04	3.7	6.48	12.86	2.792
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.4912	0.02764	Moyenne	3.4349	0.145383
Effet additif	-0.1139	0.027598	Effet additif	-0.5415	0.154755
Effet de Dominance	0.0023	0.054663	Effet de Dominance	-0.0577	0.236602

Annexe 2 : Effets des QTL par caractère et groupe de liaison.

LG V					
Mortalité			Charge Virale		
Mean	0.4297	0.061182	Mean	4.1403	0.345649
Effet additif	2.4351	0.645606	Effet additif	1.2276	0.268901
Effet de Dominance	1.4321	0.40612	Effet de Dominance	-0.8825	0.233369
LG VI					
Mortalité			Charge Virale		
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Mean	0.5223	0.0374	Mean	3.3943	0.2264
Effet additif QTL A	-2.4453	0.639199	Effet additif QTL A	0.5531	0.215465
Effet de Dominance QTL A	1.1497	0.413124	Effet de Dominance QTL A	-1.4324	0.281784
Effet additif QTL B	2.4432	0.618238	Effet additif QTL B	-0.3337	0.201707
Effet de Dominance QTL B	-1.2289	0.39518	Effet de Dominance QTL B	1.5881	0.329447
LG VII					
Mortalité			Charge Virale		
Mean	0.4605	0.069235	Mean	3.1259	0.355552
Effet additif	-0.0173	0.038157	Effet additif	-0.0872	0.195951
Effet de Dominance	-0.2795	0.053163	Effet de Dominance	-1.0263	0.273014
LG IX					
Mortalité			Charge Virale		
Mean	0.4605	0.069235	Mean	3.1259	0.355552
Effet additif	-0.0953	0.025493	Effet additif	-0.3698	0.130915
Effet de Dominance	0.1119	0.05367	Effet de Dominance	0.764	0.275616

Annexe 3 : Statistiques des composantes paternelle et maternelle par groupe de liaison par analyse de type Half-Sib

		LG V			
		Mortalité		Charge Virale	
Composante maternelle		F	Likelihood Ratio	F	Likelihood Ratio
			67.21	181.92	31.33

Composante paternelle		F	Likelihood Ratio	F	Likelihood Ratio
			60.91	166.35	59.3

		LG VII			
		Mortalité		Charge Virale	
Composante maternelle		F	Likelihood Ratio	F	Likelihood Ratio
			5.94	17.65	4.97

Composante paternelle		F	Likelihood Ratio	F	Likelihood Ratio
			8.75	25.89	6.27

		LG VI			
		Mortalité		Charge Virale	
Composante maternelle		F	Likelihood Ratio	F	Likelihood Ratio
			13.23	38.86	12.68

Composante paternelle		F	Likelihood Ratio	F	Likelihood Ratio
			29.06	83.23	24.61

		LG IX			
		Mortalité		Charge Virale	
Composante maternelle		F	Likelihood Ratio	F	Likelihood Ratio
			5.1	15.18	4.26

Composante paternelle		F	Likelihood Ratio	F	Likelihood Ratio
			5.62	16.71	7.11

Annexe 4: Statistiques de l'analyse "Tree – Large single Full-Sib family"

LGV					
Résistance			Charge Virale		
Famille Z*AC					
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
3.2	6.34	1.376	2.45	4.86	1.056
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.4378	0.042156	Moyenne	3.3704	0.201534
Effet Paternel	0.0526	0.060365	Effet Paternel	0.2715	0.27794
Effet Maternel	0.108	0.048273	Effet Maternel	0.4119	0.235042

Famille E*T					
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
24.5	45.35	9.848	9.1	17.67	3.838
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.6541	0.034759	Moyenne	4.0977	0.165972
Effet Paternel	-0.2374	0.034762	Effet Paternel	-0.7799	0.193063
Effet Maternel	-0.0899	0.053613	Effet Maternel	-0.3328	0.23292

Famille R*O					
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
110.82	165.57	35.953	71.65	116.94	25.393
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.4922	0.022501	Moyenne	3.6385	0.11419
Effet Paternel	0.3704	0.025291	Effet Paternel	1.5144	0.128346
Effet Maternel	0.09	0.043283	Effet Maternel	0.338	0.219652

LG VI					
Mortalité			Charge Virale		
Famille Z*AC					
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
4.86	9.56	2.075	3.86	7.62	1.655
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.5118	0.029534	Moyenne	3.7503	0.138549
Effet Paternel	0.1428	0.049543	Effet Paternel	0.4162	0.232413
Effet Maternel	0.0777	0.04955	Effet Maternel	0.5464	0.23245

Famille E*T					
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
14.62	27.89	6.056	17.93	33.85	7.35
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.5681	0.031322	Moyenne	3.6177	0.128698
Effet Paternel	0.0080	0.034094	Effet Paternel	1.0083	0.168404
Effet Maternel	-0.1769	0.035854	Effet Maternel	0.0079	0.168928

Famille R*O					
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
21.63	40.38	8.768	23.86	44.26	9.611
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.3569	0.034866	Moyenne	3.0086	0.161967
Effet Paternel	0.1224	0.034021	Effet Paternel	0.7953	0.158041
Effet Maternel	0.1947	0.035517	Effet Maternel	0.776	0.164991

LG VII

Résistance			Charge Virale		
Famille Z*AC					
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
2.75	5.44	1.181	1.68	3.34	0.725
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.5188	0.029885	Moyenne	3.7515	0.140078
Effet Paternel	0.1207	0.052123	Effet Paternel	0.3746	0.225287
Effet Maternel	-0.0175	0.055365	Effet Maternel	0.088	0.22852

Famille E*T					
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
6.74	13.17	2.861	5.84	11.45	2.486
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.5325	0.030342	Moyenne	3.7768	0.142374
Effet Paternel	-0.1273	0.044635	Effet Paternel	-0.683	0.209443
Effet Maternel	0.0286	0.040253	Effet Maternel	-0.1173	0.188881

Famille R*O					
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
9.68	18.74	4.07	6.44	12.61	2.739
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.5359	0.029105	Moyenne	3.8017	0.137367
Effet Paternel	-0.03	0.0299	Effet Paternel	-0.0662	0.14381
Effet Maternel	-0.1733	0.040056	Effet Maternel	-0.7014	0.196151

LG IX					
Résistance			Charge Virale		
Famille Z*AC					
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
5.76	11.31	3.156	5.44	10.69	3.32
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.4978	0.028685	Moyenne	3.654	0.134424
Effet Paternel	-0.0914	0.04312	Effet Paternel	-0.3632	0.20207
Effet Maternel	-0.1181	0.038266	Effet Maternel	-0.559	0.179325

Famille E*T					
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
1.27	2.53	0.549	3.19	6.3	1.369
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.5019	0.030166	Moyenne	3.3216	0.20822
Effet Paternel	-0.0483	0.0436	Effet Paternel	-0.3729	0.288622
Effet Maternel	-0.0528	0.047899	Effet Maternel	0.5322	0.272667

Famille R*O					
F	Likelihood Ratio	LOD	F	Likelihood Ratio	LOD
5.66	11.11	2.412	4.48	8.82	1.915
Source	Estimation	Ecart type	Source	Estimation	Ecart type
Moyenne	0.4683	0.031555	Moyenne	3.4717	0.155111
Effet Paternel	0.0443	0.088169	Effet Paternel	0.3292	0.334892
Effet Maternel	-0.1645	0.093957	Effet Maternel	-0.8679	0.375611

Résumé :

L'huître du Pacifique *Crassostrea gigas* est la principale espèce aquatique produite dans le monde (4,2 millions de tonnes pour un chiffre d'affaire de 3,5 milliards de dollars US selon la FAO, 2005). L'élevage de l'huître creuse constitue une composante de l'activité économique de la France. Cependant, depuis plus de 15 ans, les élevages subissent des épisodes de mortalité estivales (30%-60%) qui peuvent à terme mettre en péril la compétitivité de l'aquaculture européenne de mollusques. Hormis l'importance économique de *C. gigas*, qui justifie un effort de recherche, les huîtres constituent un modèle d'étude des bases physiologiques et génétiques de caractères complexes (ie croissance, reproduction et survie) fortement corrélés avec la réponse des huîtres à différentes conditions environnementales.

Une des solutions envisagées est la sélection d'individus plus résistants à la mortalité. Compte tenu des contraintes liées à la mise en place de tels schémas de sélection, le développement de marqueurs génétiques, permettant une sélection assistée par ces mêmes marqueurs, apparaît aujourd'hui comme une voie de recherche intéressante mais vise aussi à une meilleure compréhension du phénomène de mortalité estivale chez l'huître creuse par l'étude de l'architecture génétique de ce trait.

D'autre part, les connaissances sur le génome de l'huître creuse sont actuellement en plein développement (Hedgecock, Gaffney et al. 2005). Il apparaît comme particulièrement polymorphe et portant un grand nombre de gènes délétères récessifs, responsables de distorsions de ségrégation importantes (Launey et Hedgecock, 2001). Une carte de liaison basée sur une centaine de marqueurs microsatellites a été récemment publiée (Hubert et Hedgecock, 2004).

Réalisé dans le cadre du projet européen « Aquafirst » (FP6), ce sujet de doctorat a pour but de développer une cartographie nous permettant de détecter des relations entre génotype et phénotype par une approche QTL (Quantitative Trait Locus). Le trait de phénotype considéré est la mortalité estivale chez l'huître creuse *Crassostrea gigas*. Pour cela, de nouveaux marqueurs moléculaires de type SNP (Single Nucleotide Polymorphism) ont été développés pour établir une carte génétique à l'aide de l'ensemble des marqueurs disponibles (microsatellites et SNP). Ces travaux ont été menés sur des individus issus de familles d'huîtres consanguines présentant des performances contrastées en terme de survie estivale qui ont été obtenus lors du programme MOREST.