

AQUAFIRST : programme européen



TRUITE- HUITRE-BAR- DAURADE

*Combined genetic and functional genomic
approaches for stress and disease resistance marker
assisted selection in fish and shellfish.*



Francine Krieg Evry Mai 2009



Truite Arc En Ciel

Projet en 4 étapes



- Identifier et caractériser sur le plan fonctionnel des gènes impliqués dans la réponse aux situations de stress ou d'exposition à des pathogènes.
- Identifier des polymorphismes dans des gènes candidats (SNP ou μ sat) pour rechercher des associations entre ces polymorphismes et le phénotype aux caractères d'intérêt.
- Détecter des QTL pour les mêmes caractères et quelques caractères liés.
- Définir des protocoles de sélection opérationnels combinant sélection classique et MAS pour les caractères cibles

Matériel animal



1. Souches divergentes sélectionnées pendant trois générations pour la réponse cortisol sanguin induite par un stress de confinement standardisé.(T. Pottinger)
2. Clones homozygotes sensibles ou résistants à une infection virale , la SHV

1206 gènes reçus des partenaires



Stress confinement (partner C001) : 190 séquences



VHS challenge (partner C001) : 822 séquences , 322 gardés

Yersinia Challenge (partner CR02) : 194 séquences



Au final 706 gènes peuvent être étudiés

Objectifs : Construction de 3 SNPLex

Multiplex de 48 Loci - Contrainte de la méthode :

- Soumettre 10 à 20% de séquences en plus du nombre final
- Pas d'insertion délétion >6 paire de bases
- Au moins 6 paires de bases entre deux SNP (15 mieux)
- Concentrations des ADN très homogènes

Duplication récente du génome des salmonidés

Analyse d'individus homozygotes: toute « variabilité » observée sera alors impliquée à des séquences multiples. Locus non intégré dans le SNPLex

Tenant compte de ces critères , recherche de polymorphisme dans les parents des 5 familles QTL et des 2 familles de référence de la carte génétique



De la séquence au polymorphisme identifié



- Recherche des séquences répétées (mreps software)



- Recherche des séquences répétées dispersées dans le génome (salmonids repeat masker software)



- Observation de la structure des contigs : INRA bioinformatics plateforme (SIGENAE)

- Comparaison avec le génome Zebra (Ensembl software) : information sur la taille et la position des introns

- Définition des amorces avec primer3

INDIVIDUS ETUDIES

2200

- Familles QTL
- Famille de référence pour la carte génétique
- Populations commerciales
- Populations expérimentales
- Clones
- Espèce voisine : *Salmo salar*, *Salmo trutta*, hybride Clarki mykiss



Résultats

Nombre d'EST étudiées	555		
Nombre de couples d'amorces étudiés	609		
Pas d'amplification	127	20.8%	30% perdus
Amplification non spécifique	59	9.7%	
Amplification correcte, une seule bande en gel d'agarose	423	69.5%	
Etudiées	396		
Séquences multiples	139	35%	35% perdus
Ok pour étude de polymorphisme	238		
Locus monomorphe	75	31.5%	31.5% perdus

163 séquences avec du polymorphisme (30% du nombre initial)

Bilan sur l'ensemble des locus (220)

-11459 dans les introns,

-47780 dans les exons

-59 239 paires de bases séquencées

Challenge	N locus	Bp in exons	SNP/ N bp in exon	Bp in introns	SNP/ N bp in intron	% transversion	% transition	% loci with indel
Stress	62	11966	1/398	6221	1/239	50	50	28
Bacterial	47	11956	1/229	1510	1/137	48	52	36
Viral	111	23858	1/230	3728	1/150	56	44	30

Intégration dans le SNPlex de séquences avec des indel > 6 paires de base.



Allele non délété (indel en jaune): le SNP C/G est imaginé, la base G n'existe pas. La technique SNPlex détectera les individus ayant l'insertion pour au moins un des allèles.

TTTGGGTTCAATAGCTTAATCAAAGGTGAAAAGTCT [C/G] AAATGTGAGGTCCT
TTGAAAACCTGATGGATA



Allèle délété. On imagine un SNP sur la première base après la délétion. Seuls les individus porteurs de la délétion seront repérés.

GGGTTCAATAGCTTAATCAAAGGTGAG/A] GTCTTTGAAAACCTGATGG
ATATCTG



Les hétérozygotes seront déterminés par compilation des résultats

Intégration dans le SNPlex de séquences avec des SNP proches



GGGTTCAATAGCTTAATCAAAGGTGAAAAG/CTCTCAG/AGTCTTTGAAA

Même principe , 2 emplacements dans le SNPlex

1 pour l'allèle A du 2^{ème} SNP

GGGTTCAATAGCTTAATCAAAGGTGAAAAG/CTCTCAAGTCTTTGAAA

1 autre pour l'allèle G

GGGTTCAATAGCTTAATCAAAGGTGAAAAG/CTCTCAGGTCTTTGAAA



**Au final, 161 séquences représentant 120 gènes
ont été proposées pour construire les SNPlex.**

Séquences retenues

161 séquences



142 retenues



142 séquences = 120 locus



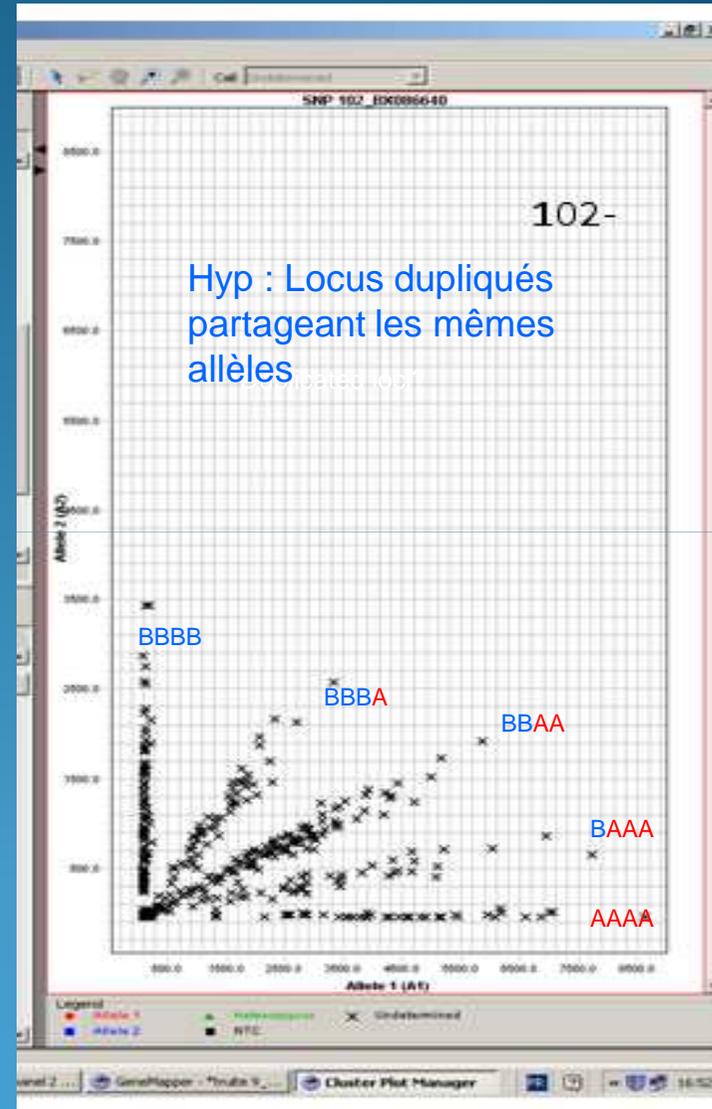
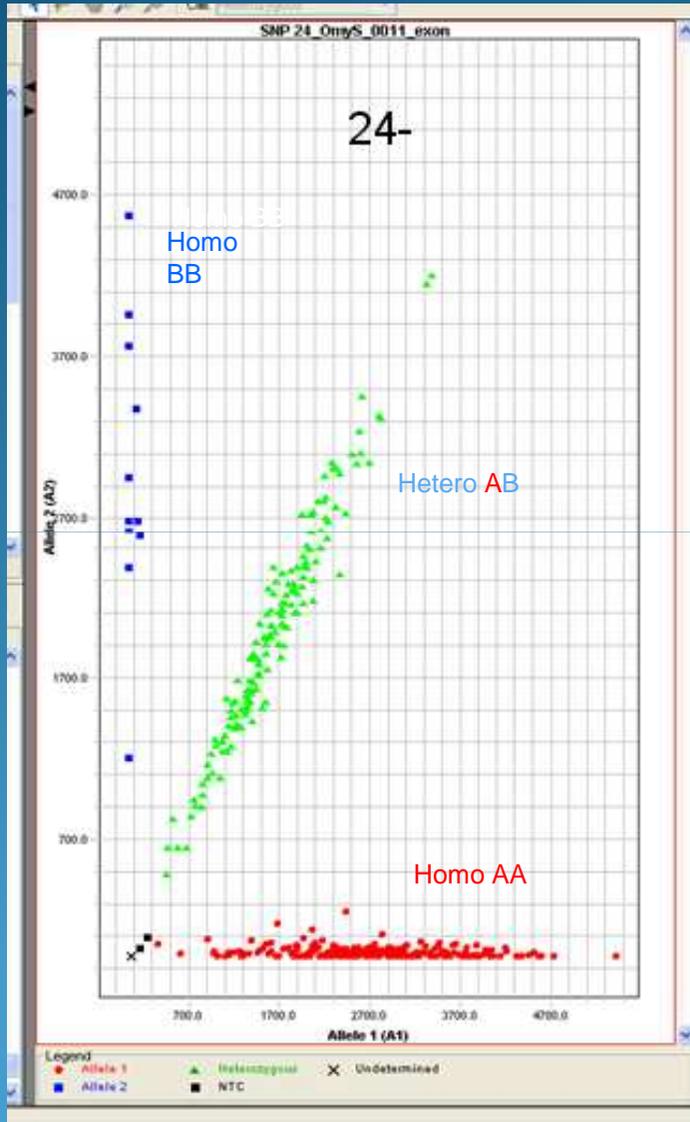
- 6 locus mis deux fois pour cartographie et recherche de QTL



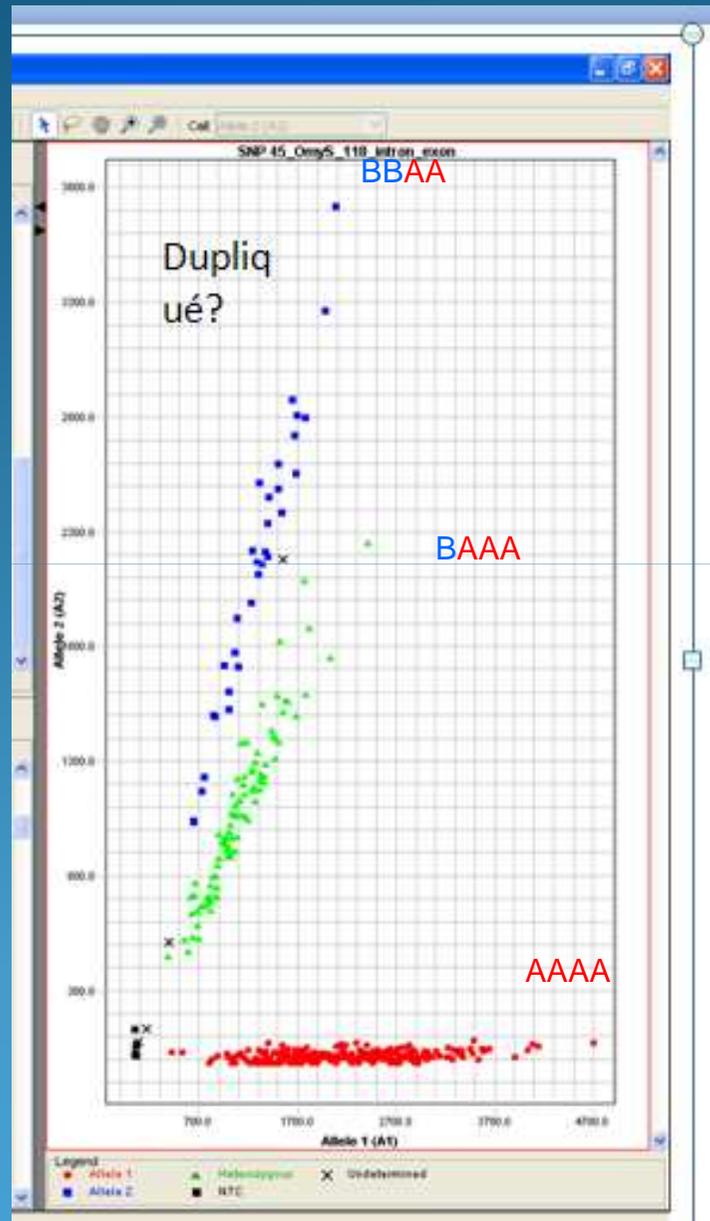
- 7 locus mis en deux fois pour indel >6 paires de base

- 9 locus mis en deux fois pour SNP proches

Images SNPlex

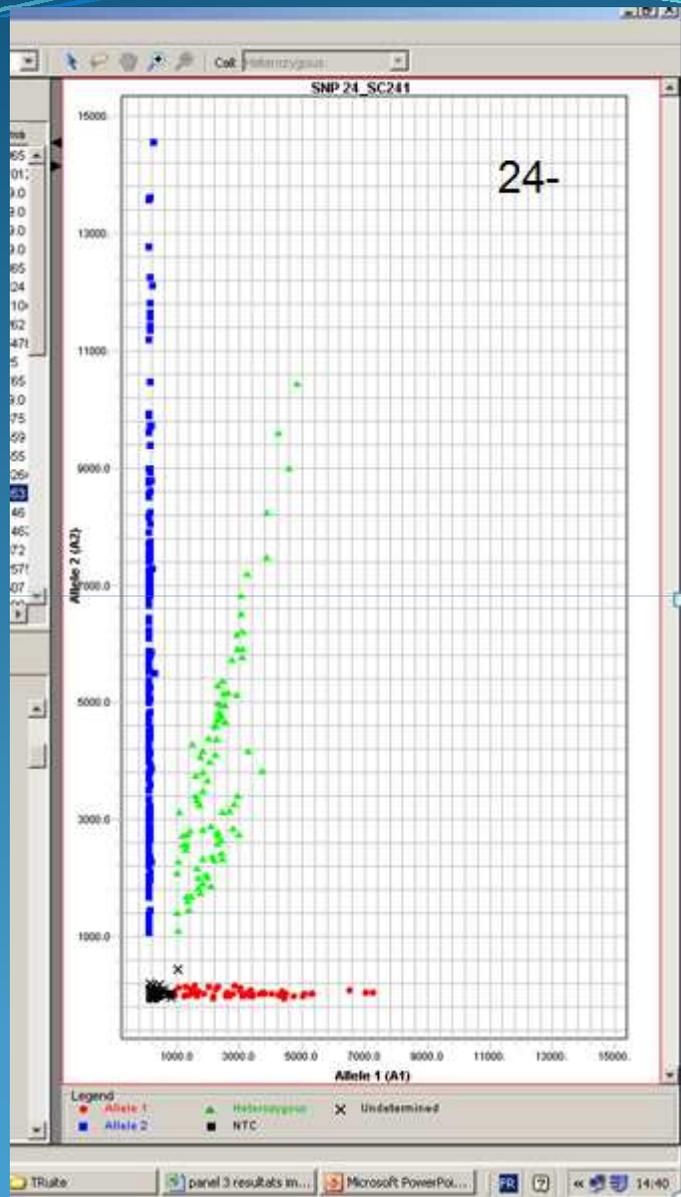


Images SNPlex

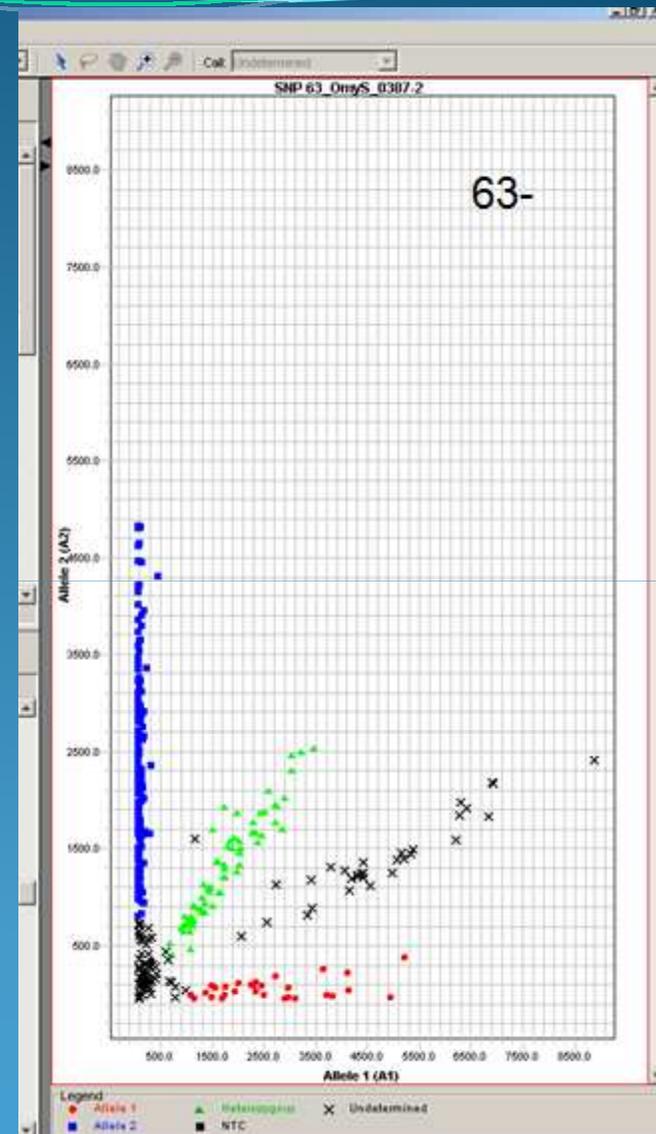


HYP : Locus dupliqué - 1 fixe ,
1 variable

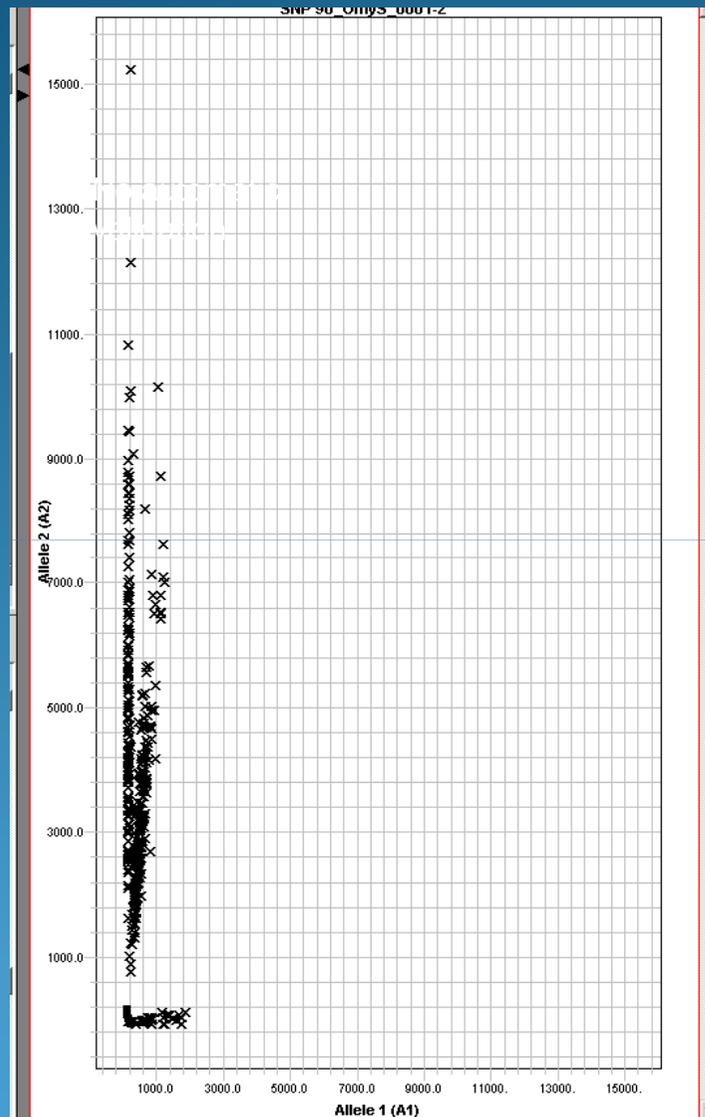
Images SNPlex



?



Images SNPlex



Validation manuelle
difficile



Bilan SNPlex

Deux séries d'analyse

Novembre 2007.

Panel 1 les 3 premières plaques 384 : Résultats satisfaisants. Sur les deux premières plaques 80 à 90% de validation automatique, sur la troisième un peu moins (60%) mais concentration des ADN plus faible

Juin 2008

Panel1 : les trois dernières plaques

Panel 2 et 3 : l'ensemble des plaques

En moyenne deux locus par plaque ont été validés automatiquement par le logiciel , soit une validation manuelle pour la quasi totalité



15 locus n'ont pu être analysés (12.5%)



6 Hyp : séquences multiples



3 : valeurs de pics trop faibles

4 : indel mis en deux fois , un des deux trop faible



2 : SNP proches, + 3 autres suspects

85 locus cartographiés

20 locus monomorphe dans les familles de la carte
mais possibilité avec les familles QTL



GL	N locus	GL	N locus	GL	N locus
GL1	6	GL10	4	GL20	3
GL2A+29	10	GL11	2	GL21	4
GL2B	1	GL12	4	GL22	8
GL3	7	GL13	2	GL23	2
GL4	2	GL14	6	GL24	6
GL5	4	GL15	0	GL25	1
GL6	2	GL16	6	GL26	2
GL7	3	GL17	1	GL27	5
GL8	5	GL18	1	GL30	2
GL9	7	GL19	5	GL31	6

CONCLUSION : SNPlex = bilan mitigé

Positif :

Rapidité si validation automatique
Une fois les SNP validés possibilité d'utilisation pour la réassignation de parenté (intéressants pour les producteurs)

Problèmes liés à notre matériel animal

Duplication: Eviction de certains gènes , malgré cela apparition d'accrochages multiples
Taux d'insertion-délétion élevé (2 emplacement au lieu d'1, temps d'analyse augmenté)
Pour les SNP proches difficulté de lecture , accroche moins spécifiques des amorces?

Problèmes liés à ???

Pourquoi tant de validation manuelle lors de la deuxième série d'analyse.
Validation manuelle malgré des valeurs de pics tout à fait correcte, et parfois validation automatique de valeur aberrantes

MERCI



Aquafirst Partenaires

P. Prunet (scribe , INRA)
T. Pottinger NERC, UK
C. Secombes UNI ABDN, UK

INRA : GABI équipe GenAqua

E. Quillet
C.Hervet
K.Canale-Tabet
A.Launay
S.Mezi (CDD)
L.Laffont (CDD)

CNG : EPGV-INRA_UR1279

D. Brunel
A. Bérard
A. Chauveau

