



UNIVERSITA' DEGLI STUDI DI PADOVA

Sede amministrativa: Università degli Studi di Padova

**DIPARTIMENTO DI SANITA' PUBBLICA, PATOLOGIA
COMPARATA E IGIENE VETERINARIA**

**SCUOLA DI DOTTORATO IN SCIENZE VETERINARIE
INDIRIZZO DI SANITA' PUBBLICA E PATOLOGIA COMPARATA**

Ciclo XXII

APPLICAZIONE DELLA GENOMICA FUNZIONALE ALLO STUDIO DI SPECIE DI INTERESSE IN ACQUACOLTURA

Direttore della scuola: Ch.mo Prof. Massimo MORGANTE

Coordinatore di indirizzo: Ch.mo Prof. Mauro DACASTO

Supervisore: Prof. Luca BARGELLONI

Dottoranda: Serena FERRARESSO

31 GENNAIO 2010

Riassunto

Negli ultimi 10-15 anni, la produzione di specie marine in acquacoltura è cresciuta molto rapidamente; diverse limitazioni, tuttavia, non sono ancora state superate (per es. elevata mortalità larvale, suscettibilità a stress e patologie, sviluppo di malformazioni scheletriche). La genomica funzionale può fornire gli strumenti per una migliore conoscenza dei meccanismi molecolari responsabili di tratti produttivi di forte interesse economico nelle specie allevate.

Nel presente studio, numerosi sforzi sono stati investiti allo scopo di sviluppare e sfruttare strumenti di genomica funzionale per lo studio di due specie di primaria importanza per l'acquacoltura europea, l'orata (*Sparus aurata*) e il branzino (*Dicentrarchus labrax*); di seguito viene riportato lo sviluppo di due piattaforme *microarray* a oligonucleotidi per entrambe le specie.

Tutti i trascritti disponibili di orata sono stati utilizzati per la sintesi di due sonde a oligonucleotidi (60meri) per ciascun trascritto utilizzando la tecnologia Agilent *SurePrint*TM. La riproducibilità e l'accuratezza della piattaforma *microarray* sono state quindi testate in due stadi precoci di sviluppo di orata: due e quattro giorni dalla schiusa. La correlazione tra le repliche tecniche si è sempre rivelata elevata ($r > 0,99$) e forte correlazione è stata osservata anche tra le coppie di sonde per ciascun trascritto. La *Real-time* RT-PCR è stata quindi utilizzata per cross-validare i dati di espressione del *microarray*; anche in questo caso è stata ottenuta una correlazione statisticamente significativa per ciascun gene *target* in tutte le repliche biologiche.

Per il branzino, il disegno delle sonde è stato effettuato per 19.035 trascritti *target*; l'elevata affidabilità e riproducibilità della piattaforma è stata confermata tramite *Real-time* RT-PCR e analisi delle repliche biologiche. Il *microarray* è stato quindi impiegato per analizzare i profili di espressione genica in mandibole e teste di esemplari di branzino affetti da prognatismo, una malformazione scheletrica che ne compromette pesantemente la produzione e la commerciabilità. Analisi statistiche hanno permesso di identificare 242 trascritti sottoespressi negli esemplari prognati rispetto ai controlli; una annotazione funzionale di questi fornisce nuove informazioni sui potenziali meccanismi coinvolti nello sviluppo di questa malformazione.

Nel presente studio viene anche descritta l'identificazione e la caratterizzazione, in branzino, dei recettori *Novel Immune-type Receptor* (NITR). In totale sono stati identificati 27 geni e 3 pseudogeni, organizzati in *cluster* in una regione genomica di circa 270 kb. I membri di questo *cluster* sono caratterizzati da una elevata variabilità; i NITR, infatti, mantengono tre organizzazioni genomiche principali, che appaiono conservate in diverse specie di teleostei, oltre a forme nuove che coinvolgono processi di perdita di introni, delezioni di esoni o loro nuova acquisizione. Analisi comparative tra tutti i geni NITR finora noti sembrano suggerire che questi seguano il modello evolutivo noto col nome di "*birth-and-death*", nel quale eventi di duplicazione, assieme alle perdite/acquisizioni di singoli membri, contribuiscono alla rapida evoluzione di diverse famiglie geniche.

Summary

In the last 10-15 years, the production of farmed marine fish species has increased rapidly, mainly as a consequence of improved breeding and husbandry methods and technologies. However, severe production bottlenecks, such as high larval mortality, skeletal malformations, susceptibility to stress and disease, remain to be solved. The application of genomic technologies offers much promise to rapidly increase our knowledge on molecular mechanisms underlying key productive traits in farmed species.

In the present study, many efforts were invested in order to provide genomic tools for the study of two fish species of great importance for the European aquaculture; the gilthead seabream (*Sparus aurata*) and the European sea bass (*Dicentrarchus labrax*). To this end, we report here on the development of two oligo DNA microarray for both *S. aurata* and *D. labrax*.

All the available ESTs from the gilthead sea bream were used to design two longer (60mer) probes for each transcript using Agilent SurePrint™ technology. This microarray platform was then validated to assess its reproducibility and accuracy on two early stages of gilthead sea bream development, respectively one-day and four days old larvae. Correlation between technical replicates was always > 0.99, with strong positive correlation between paired probes. Cross-validation of microarray data was then carried out using quantitative (q)RT-PCR; a statistically significant positive correlation was obtained comparing expression levels for each target gene across all biological replicates.

For the European sea bass, probe design was positively completed for 19,035 target transcripts; high levels of reliability, and reproducibility of the microarray platform were confirmed by qPCR validation and biological replicates analysis. The oligo microarray was then applied to profile gene expression in mandibles and whole-heads of fish affected by prognathism, a skeletal malformation that strongly affect sea bass production. Statistical analysis identified 242 transcripts that are significantly down-regulated in deformed individuals compared to normal fish. Functional annotation of significant genes gives new insights on putative mechanisms involved on mandibular prognathism development.

In this study we describe also the identification and characterisation of a NITR gene complex *D. labrax*, encoded in a single genomic region spanning 270 kilo bases (kb). In total, 27 NITR genes and 3 pseudogenes, organized in a tandemly arrayed cluster, were identified. Comparison of the predicted peptide structure across *D. labrax* NITRs shows considerable variations; these genes maintain the three major genomic organizations that appear to be essentially conserved among fish species along with new features presumably involving processes of intron loss, exon deletion, and acquisition on new exons. Comparative analyses of all published NITR genes suggest that these receptors follow birth-and-death model of gene evolution in which duplication events together with lineage-specific gain and loss of individual members contributed to the rapid diversification of individual gene families.

PUBBLICAZIONI RELATIVE AL PROGETTO DI RICERCA

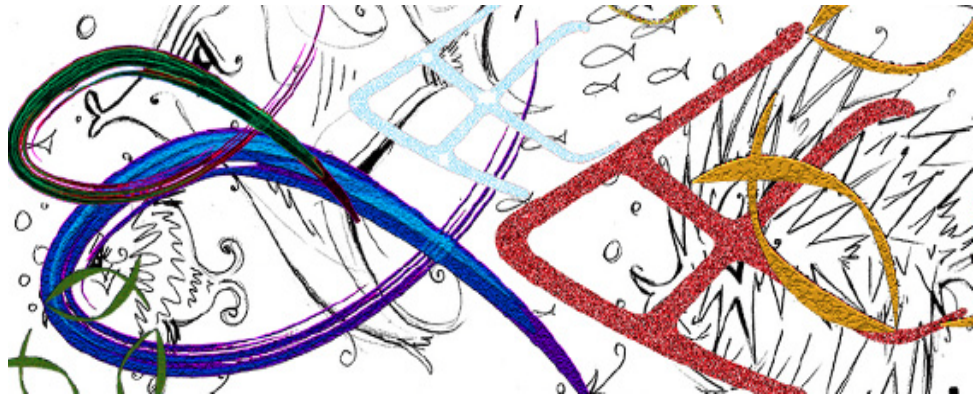
Ferraresso S, Kuhl H, Milan M, Ritchie DW, Secombes CJ, Reinhardt R, Bargelloni L: **Identification and characterisation of a novel immune-type receptor (NITR) gene cluster in the European sea bass, *Dicentrarchus labrax*, reveals recurrent gene expansion and diversification by positive selection.** *Immunogenetics* 2009, **61**:773-88.

Ferraresso S, Vitulo N, Mininni AN, Romualdi C, Cardazzo B, Negrisolò E, Reinhardt R, Canario AVM, Patarnello T, Bargelloni L: **Development and validation of a gene expression oligo microarray for the gilthead sea bream (*Sparus aurata*).** *BMC Genomics* 2008, **9**:580.

2. IDENTIFICAZIONE E CARATTERIZZAZIONE DEI NOVEL IMMUNE-TYPE RECEPTOR NEL BRANZINO	31
2.1 Introduzione	33
2.2 Materiali e metodi	39
2.2.1 Campioni biologici ed estrazione dell'RNA	39
2.2.2 Sequenziamento dei cloni BAC di <i>D. labrax</i>	39
2.2.3 Identificazione <i>in silico</i> dei geni NITR	40
2.2.4 Amplificazione del cDNA e sequenziamento di trascritti <i>full-length</i>	41
2.2.5 Analisi filogenetiche	42
2.2.6 Analisi di pressione selettiva	42
2.2.7 Analisi di genomica comparativa	43
2.2.8 <i>Modelling</i> comparativo del dominio V	44
2.2.9 Analisi di espressione dei geni NITR in <i>D. labrax</i>	44
2.3 Risultati	47
2.3.1 Identificazione di un cluster genico di NITR in <i>D. labrax</i>	47
2.3.2 I geni NITR condividono elementi conservati	49
2.3.3 Il <i>cluster</i> genico dei NITR è costituito da diverse famiglie di recettori	50
2.3.4 I NITR seguono il modello evolutivo "<i>birth-and-death</i>"	53
2.3.5 I geni NITR sono soggetti a selezione positiva	55
2.3.6 I NITR di branzino mostrano relazioni sinteniche con altre specie di teleostei	59
2.3.7 Analisi dell'espressione genica dei NITR di <i>D. labrax</i>	60
2.4 Discussione	63

3. SVILUPPO E VALIDAZIONE DI UNA PIATTAFORMA MICROARRAY PER SPARUS AURATA	71
3.1 Introduzione	73
3.1.1 La tecnologia Agilent	75
3.2 Materiali e metodi	77
3.2.1 Campioni biologici ed estrazione dell'RNA	77
3.2.2 Costruzione del <i>database</i> per <i>Sparus aurata</i> e disegno delle sonde	77
3.2.3 Esperimenti di <i>microarray</i> e analisi dei dati	80
3.2.4 Validazione dei dati di espressione mediante <i>Real-time</i> RT-PCR	82
3.3 Risultati	83
3.3.1 Costruzione del <i>database</i> SAPD	83
3.3.2 Analisi dei dati di <i>microarray</i>	84
3.3.2.1 <i>Valutazione delle performance della piattaforma <i>microarray</i></i>	84
3.3.2.2 <i>Analisi dei geni differenzialmente espressi</i>	88
3.3.3 Analisi di espressione tramite <i>Real-time</i> RT-PCR	89
3.4 Discussione	93

4. SVILUPPO DI UNA PIATTAFORMA <i>MICROARRAY</i> PER LO STUDIO DEL PROGNATISMO IN <i>DICENTRARCHUS LABRAX</i>	101
4.1 Introduzione	103
4.2 Materiali e metodi	107
4.2.1 Campioni biologici ed estrazione dell'RNA	107
4.2.2 DLDP <i>database</i> e annotazione dei trascritti di <i>D. labrax</i>	108
4.2.3 Disegno delle sonde ed esperimenti di <i>microarray</i>	109
4.2.4 Acquisizione ed analisi dei dati sperimentali	110
4.2.5 Annotazione funzionale dei trascritti di <i>D. labrax</i>	111
4.2.6 Validazione dei dati di espressione mediante <i>Real-time</i> RT-PCR	113
4.3 Risultati	115
4.3.1 <i>Database</i> DLDP	115
4.3.2 Valutazione della qualità della piattaforma <i>microarray</i>	117
4.3.3 Analisi dei dati di espressione genica	120
4.3.4 Annotazione funzionale dei geni differenzialmente espressi	121
4.3.5 Validazione mediante <i>Real-time</i> RT-PCR	123
4.4 Discussione	127
5. CONCLUSIONI GENERALI	135
<i>Bibliografia</i>	141
<i>Appendici</i>	163



1. INTRODUZIONE GENERALE

1.1 Acquacoltura dal passato ad oggi

L'acquacoltura è una realtà presente nella storia dell'uomo fin dall'antichità. Reperti archeologici hanno infatti rivelato testimonianze circa l'utilizzo di tecniche per allevare pesci in zone circoscritte già nel 2500 a.C. Il bisogno di mantenere costante nel tempo la disponibilità di specie acquatiche per scopi alimentari, liberandosi dalla incertezza delle riserve naturali, ha portato numerose civiltà a sviluppare tecniche di allevamento ittico. Sebbene praticata già dai Romani, l'acquacoltura ha registrato in Europa una forte diffusione a partire dal Medioevo a causa della scarsità e dell'elevato prezzo del pescato lontano dalle coste e i grandi corsi d'acqua. Nella seconda metà del secolo scorso l'itticoltura è stata il settore delle produzioni animali di più rapido sviluppo; da una produzione globale inferiore ad 1 milione di tonnellate nei primi anni '50, ha raggiunto i 51,7 milioni di tonnellate nel 2006, mantenendo un tasso di crescita annuo che sfiora il 7% (FAO, 2008), che supera quello della popolazione mondiale stessa. La produzione ittica sta giocando un ruolo sempre maggiore nel soddisfare la richiesta di consumo da parte dell'uomo di pesce e prodotti della pesca; il contributo medio di questo settore alla disponibilità di pescato è infatti passato dal 14% del 1986 al 47% del 2006, con un apporto *pro capite* cresciuto dagli 0.7 kg nel 1970 ai 7.8 kg del 2006.

Secondo quanto riportato in “*World review of fisheries and aquaculture*”, tra il 1995 e il 2007, la produzione globale il pesce allevato è all'incirca triplicato in volume raggiungendo quest'anno un importante traguardo: quello del 50% della produzione di tutto il pesce consumato nel mondo (Naylor *et al.* 2009).

I maggiori produttori sono le regioni asiatiche e del Pacifico che coprono l'89% ed il 77% della produzione mondiale rispettivamente in termini di quantità e di valore commerciale; questo è dovuto principalmente alla Cina che da sola detiene il 67% della produzione globale di pescato. Secondo lo “*The State of World Fisheries and Aquaculture*”, il massimo potenziale della pesca di cattura è ormai stato raggiunto, attualmente non si prevede un ulteriore aumento produttivo del comparto pesca, grazie anche alle normative internazionali contro l'impoverimento dei mari. Questa situazione rappresenta una ulteriore *chance* di sviluppo per l'allevamento ittico di numerose specie tradizionalmente catturate in contesto europeo e americano.

1.1.1 Stato dell'arte della piscicoltura in Europa

La produzione del settore dell'acquacoltura in Europa ha raggiunto nel 2006 oltre 1.415.000 tonnellate (t) di pescato. Tra i 25 stati membri dell'Unione Europea, i principali produttori risultano essere il Regno Unito (145.739 t), la Grecia (83.000 t), l'Italia (71.900 t), la Spagna (65.515 t) seguita dalla Francia (49.900 t).

Considerando l'andamento delle ultime 2 decadi, in Europa il settore della pesca è andato incontro ad un declino passando dai 9 milioni di tonnellate del 1989 ai circa 7 milioni del 2005, mentre l'acquacoltura ha mantenuto un costante sebbene lieve incremento. Le specie principalmente allevate sono la trota (*Oncorhynchus mykiss*), il salmone (*Salmo salar*) seguiti da specie eurialine come l'orata (*Sparus aurata*) e branzino (*Dicentrarchus labrax*) e dall'anguilla (*Anguilla anguilla*) (Melotti *et al.* 2009). I principali produttori di salmone si confermano essere i paesi nord europei con al primo posto la Norvegia (produttrice di oltre 626.000 t) mentre, per quanto riguarda la produzione di trota, l'Italia si posiziona al primo posto tra i produttori europei (con 39.600 t). La produzione di specie eurialine è concentrata principalmente nell'area del Mediterraneo, le specie principalmente allevate sono senza dubbio branzino e orata con una produzione rispettivamente di oltre 84.000 t e 107.000 t nel 2006. Per entrambe le specie, la Grecia è il principale produttore europeo mentre l'Italia si classifica quarto paese in Europa per la produzione sia di orate che di branzini mantenendo un tasso di crescita sensibile ma tuttavia inferiore a quello di Turchia e Spagna.

La produzione di anguille in Europa assume dimensioni decisamente inferiori con una resa complessiva di poco più di 8.000 t, la maggior parte proveniente dall'Olanda (4.200 t), Danimarca (1.750 t) e Italia (1.700 t). In controtendenza rispetto alle altre colture, l'allevamento di questa specie è l'unico che ha mostrato un costante decremento dagli anni '70 a causa della continua riduzione di esemplari giovani provenienti dall'Atlantico.

1.2 La piscicoltura nel Mediterraneo: orata e branzino

Sino alla fine degli anni '70, la piscicoltura nel Mediterraneo era basata esclusivamente sulla raccolta in natura di giovanili per l'allevamento in sistemi estensivi e semintensivi. La produzione a carattere estensivo (come ad esempio la Vallicoltura), è stato infatti il sistema di allevamento predominante anche in Italia per lungo periodo.

Il branzino e l'orata, sono state le prime specie pregiate eurialine ad essere state riprodotte in condizioni controllate ed allevate in sistemi intensivi.

La vallicoltura tradizionale ha visto, infatti, una progressiva riduzione del potenziale produttivo a partire dagli anni '70-'80 in seguito alla diminuzione di soggetti selvatici deputati alla semina. Il forte aumento della richiesta di avannotti, non più soddisfatto dalla pesca, stimolò pertanto la messa a punto di tecniche per la produzione ed il trasporto di uova, larve e giovani. Piccole unità produttive comparvero nei primi anni '80 in Italia ed in Francia, garantendo una fonte alternativa di approvvigionamento per gli impianti nel bacino del Mediterraneo. I primi programmi di allevamento sono iniziati negli anni '90 dapprima in Francia e Israele seguiti poi da Grecia, Spagna e Italia (Chavanne *et al.* 2009). La standardizzazione delle tecniche per la produzione ed il trasporto di uova, larve e giovanili per diverse specie ittiche ha così permesso l'industrializzazione dell'acquacoltura.

Ad oggi, la produzione mondiale complessiva di branzino (*D. labrax*), derivante da stock selvatici e impianti di acquacoltura, si è attestata nel quinquennio 2000-2005 al valore medio di 60.000 tonnellate annue, sfiorando le 88.000 tonnellate nel 2006. La produzione italiana copre un volume di circa 11.000 t annue, di cui il 70% derivante da allevamenti. Per quanto riguarda l'orata la produzione è ancora maggiore raggiungendo le 107.000 t nel 2006, anche in questo caso il contributo italiano è significativo e si attesta attorno al 17%.

1.2.1 Orata (*Sparus aurata*)

Sparus aurata Linnaeus, 1758

Classe: Actinopterygii

Ordine: Perciformes

Famiglia: Sparidae



L'orata, *Sparus aurata* (Linnaeus 1758), è un teleosteo marino della famiglia *Sparidae* che comprende più di 100 specie. Gli Sparidi sono ampiamente diffusi in acque temperate e tropicali, con un massimo di diversità di specie nelle regioni del Nord-Est Atlantico e del Mediterraneo, e rappresentano un elemento chiave nell'ecosistema costiero marino. Molte delle specie della famiglia degli sparidi rivestono un ruolo di rilievo nell'ambito dell'economia della pesca e dell'acquacoltura, visto l'elevato valore commerciale conferito da qualità alimentari riconosciute a livello mondiale. Diffusa nel Mediterraneo e nell'Atlantico orientale, dalla Gran Bretagna al Senegal, è possibile trovare questa specie anche presso l'isola di Capo Verde e raramente nel Mar Nero. E' una specie eurialina e quindi tollera variazioni anche notevoli del grado di salinità pur non adattandosi ad ambienti acquatici del tutto dolci. Resiste a temperature che possono raggiungere anche i 32-33°C, mentre risente fortemente del freddo non sopravvivendo a temperature inferiori ai 5°C (Ravagnan, 1992).

Come gli altri sparidi, l'orata ha un corpo alto, ovale e massiccio, compresso lateralmente e rivestito di spesse scaglie. La mascella superiore è lievemente più lunga di quella inferiore; le labbra sono carnose ed evidenti; ha da quattro a sei canini molto robusti nella parte anteriore di ciascuna mascella, seguiti da quattro o cinque file di denti molariformi nella mascella superiore e da tre o quattro file in quella inferiore (Pepitone *et al.*, 1995). La pinna caudale è biloba.

La colorazione è grigia con riflessi azzurro dorati sul dorso e sono presenti cinque strie verticali nere e quattro grigie sui fianchi. La lunghezza media è di circa 30-35 cm, con un massimo di 70 cm. Il suo peso si aggira sui 10 kg circa, tuttavia la taglia commerciale ottimale è attorno ai 350-400 g. Si nutre principalmente di molluschi. L'orata è moderatamente gregaria, vive lungo la fascia costiera, in prevalenza su fondali sia sabbiosi sia rocciosi, generalmente da una profondità di 1-2 m sino a 30-40 m.

Durante il periodo primaverile ed estivo compie frequenti migrazioni trofiche in ambienti lagunari ed estuarini, dove, in autunno, in corrispondenza del periodo riproduttivo, viene catturata durante la risalita a mare (Crosetti *et al.*, 2001). La riproduzione avviene in autunno, in condizioni di cattività la deposizione delle uova avviene generalmente a intervalli di 24 ore per un periodo di tre mesi (generalmente da ottobre a dicembre).

L'orata è una specie ermafrodita proterandrica; sviluppa prima gli elementi maschili, poi, entro il secondo anno di età, quelli femminili. La maturità sessuale viene raggiunta nei maschi all'età di 1-2 anni (20-30 cm) e nelle femmine a 2-3 anni (33-40cm). Il percorso di inversione da maschio a femmina tuttavia non è del tutto chiaro, l'evoluzione di ciascun esemplare non è facile da predire e sembra che non tutti i maschi di orata siano soggetti allo stesso modo al fenomeno dell' inversione sessuale (in alcuni sembra assente). I meccanismi che sottendono questo fenomeno sono tuttora ignoti anche se è probabile questo avvenga in relazione a fattori genetici ambientali e sociali (Meiri *et al.*, 2002) che possono essere sfruttati per ottimizzare il rapporto maschi/femmine dei riproduttori in acquacoltura.

1.2.2 Branzino (*Dicentrarchus labrax*)

Dicentrarchus labrax Linnaeus, 1758

Classe: Actinopterygii

Ordine: Perciformes

Famiglia: Moronidae



Il branzino o spigola, *Dicentrarchus labrax* (Linnaeus, 1758), è un teleosteo marino appartenente alla famiglia *Moronidae*. In natura vive nell'Atlantico dal Senegal alla Norvegia, in tutto il Mediterraneo e nel Mar Nero. *D. labrax* è una specie euritermica (5-28 °C) ed eurialina, sopporta infatti salinità dal 3‰ al 38‰, capace di vivere sia in mare aperto che in acque salmastre o dolci. Si sposta da uno all'altro di questi *habitat* spesso seguendo il flusso delle maree. In mare è presente abitualmente lungo le coste ad una profondità compresa tra i 10 e i 100 metri; in acque salmastre si incontra con maggior frequenza nelle lagune e spesso s'inoltra all'interno dei porti o in vicinanza di scarichi urbani. In acqua dolce preferisce gli ambienti dei fiumi a lenta corrente, dei quali risale il corso anche per molti chilometri.

Considerata specie di pregio, il branzino ricopre un ruolo di primaria importanza nell'acquacoltura europea ed in particolar modo mediterranea.

Il corpo è allungato e leggermente compresso, coperto da scaglie ctenoidi di medie dimensioni. La testa è grande, con bocca terminale ampia, moderatamente protrattile e mandibola prominente. Il bordo mascellare si estende posteriormente fino al margine anteriore dell'occhio. Numerosi denti sottili e conici sono disposti in più serie su mascelle, vomere e palato. Gli occhi sono relativamente grandi. Sul dorso sono presenti due pinne dorsali non contigue. Il colore di fondo della livrea è grigio o grigio-verdastro sul dorso, i fianchi sono grigi con riflessi argentati ed il ventre è bianco. La colorazione

è variabile a seconda dell'ambiente, nelle lagune ed in allevamento appare verdastra, mentre l'habitus marino è grigio argenteo.

Negli immaturi il dorso e i fianchi sono cosparsi di piccole macchie scure. Le pinne sono chiare o grigie, con sfumature nerastre.

A tre anni le spigole superano di norma i 500 g e a quattro anni possono già toccare 1 kg di peso. Le dimensioni medie dei soggetti pescati si aggirano intorno ai 45cm e i 4 kg, ma un individuo adulto può raggiungere e superare il metro e venti di lunghezza e i 12 kg di peso (Porcellotti, 2005).

I branzini adulti vivono generalmente a coppie o isolati, mentre i soggetti immaturi ed i sub-adulti sono gregari. I branchi di giovani compiono in primavera delle migrazioni trofiche dalle acque costiere a quelle salmastre richiamati dal rapido riscaldamento delle lagune. Da gennaio a febbraio i giovani adulti compiono invece migrazioni riproduttive dalle acque dolci e salmastre al mare. Generalmente il branzino ha il suo picco di maggiore attività durante le ore notturne o crepuscolari, ma in particolari condizioni si muove anche durante il giorno, specialmente durante le mareggiate quando le acque sono torbide. Essendo una specie spiccatamente carnivora, in ambiente marino preda altri pesci (piccole anguille, cefali, latterini), cefalopodi, crostacei, molluschi e vermi policheti. Nelle acque dolci e salmastre, oltre a pesci ed invertebrati, può includere nella sua dieta micromammiferi.

Il branzino si riproduce in piccoli gruppi in acque costiere con salinità superiore al 30‰. Presenta un ciclo riproduttivo stagionale, la frega avviene una volta l'anno nei mesi invernali: nel Mediterraneo orientale da gennaio a marzo, sulle coste britanniche da marzo a giugno. La prima stagione riproduttiva avviene tra il 2° e 4° anno di vita nella popolazione del Mediterraneo orientale; nell'Atlantico la maturità viene, invece, raggiunta più tardivamente: i maschi tra il 4°-7° anno, le femmine tra 5°-8° (Chavanne *et al.*, 2009; Agbayani, 1999). La strategia riproduttiva, tipica dei teleostei, si basa sulla

fecondazione esterna di un numero elevato di uova pelagiche iponeustoniche, non sorvegliate ma disperse dalle correnti, rilasciate in concomitanza dei periodi di proliferazione planctonica, per garantire alla prole nella fase larvale zooplanctofaga, disponibilità trofica sufficiente.

1.3 Problematiche economiche legate all'acquacoltura

1.3.1 Le deformità

Le malformazioni scheletriche sono considerate tutt'oggi un problema importante nell'acquacoltura intensiva. Nonostante i progressi nel campo della nutrizione, il miglioramento dei metodi di incubazione e svezzamento delle larve, e lo scrupoloso monitoraggio delle patologie abbiano aumentato la sopravvivenza e la qualità degli avannotti, le deformità continuano ad essere presenti.

Il manifestarsi di anomalie scheletriche comporta conseguenze economiche significative per gli allevatori, poiché, sebbene non sempre riducano le prestazioni in termini di accrescimento e di resistenza a condizioni di stress (manipolazione, anestesia, sovraffollamento, infezioni), hanno forti ripercussioni sui costi produttivi e sul valore commerciale dei soggetti allevati. Le avannotterie si impegnano, quindi, a garantire alla vendita un tasso di soggetti malformati inferiore al 4-5% applicando, se necessario, una selezione manuale degli avannotti. Questo comporta un impegno di manodopera non trascurabile abbinato alla perdita economica data dai soggetti scartati nonché un rilevante problema di benessere animale (Boglione *et al.* 2001).

Molteplici fattori sono stati suggeriti come possibili cause di anomalie nello sviluppo: fattori ambientali (come per esempio la densità degli embrioni durante l'incubazione, lo stress dei riproduttori durante la vitellogenesi, la presenza di contaminanti nell'acqua, eccessiva turbolenza nelle vasche di schiusa), nutrizionali (deficienza di acido ascorbico, triptofano, fosfolipidi ed eccesso di vitamina D e A nell'alimento) e tossici (contaminanti industriali,

erbicidi e metalli pesanti). Solo di recente è stata evidenziata l'importanza di fattori genetici, finora considerati generalmente marginali, nell'eziologia delle deformità.

1.3.1.1 Il prognatismo

Il prognatismo mandibolare o protrusione del labbro inferiore, è una malformazione dello sviluppo che determina un fenotipo facciale particolarmente distintivo. Uno degli esempi più noti di prognatismo è la famiglia reale degli Asburgo, dove questa malformazione è stata riscontrata per diverse generazioni (Chudley, 1998).

Oltre che nell'uomo, questa malformazione è stata riscontrata in specie animali appartenenti a diverse classi (*reptilia*, *aves*, *mammalia*, *amphibia* e negli *actinopterygii*) nonché alcune razze di cani brachicefali (Ritchie *et al.*, 1997).

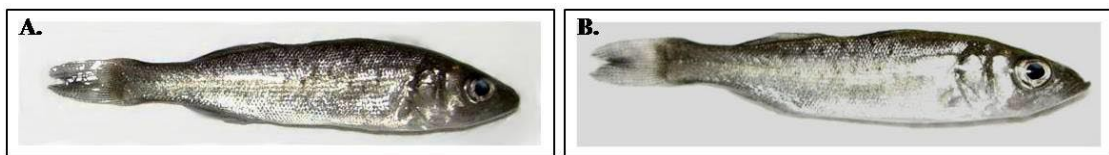


Figura 1.1: Esempio di branzino normale (A.) e prognato (B.) a 90 giorni dalla schiusa.

Il prognatismo viene definito come un'alterazione nell'accrescimento dei segmenti ossei di mandibola e mascella, che determina un raffronto anomalo tra le superfici dentarie con conseguente malocclusione, e problemi masticatori.

È possibile distinguere nell'uomo un prognatismo mascellare o enognatismo (dall'etimologia *pro* davanti e *gnathos* - *ganathos* mascella), quando è presente un eccessivo sviluppo dell'osso mascellare ed incisivo in concomitanza con un corpo mandibolare poco sviluppato; ed un prognatismo mandibolare o progenismo legato al maggior accrescimento del corpo mandibolare rispetto all'arcata mascellare. Nel primo caso la mascella risulterà sporgente ed il mento sfuggente, nel secondo sarà la mandibola a sporgere (detto comunemente labbro all'Asburgo) (Raso, 1981).

Nei pesci teleostei le strutture anatomiche che delimitano il cavo orale derivano dal primo arco viscerale che durante l'organogenesi origina la porzione craniale dello

splanocranio dividendosi in cartilagine palato-quadrata e cartilagine di Meckel, e da porzioni del secondo arco viscerale o ioideo che originano lo iosimpectico. Dall'abbozzo della cartilagine di Meckel si sviluppa anteriormente, tramite ossificazione intramembranosa paracondrale, l'osso dentale mentre posteriormente si accresce, per ossificazione endocondrale, l'osso angolare, articolato tramite il quadrato al neurocranio. Queste strutture rappresentano la base ossea della mandibola.

È possibile ipotizzare che i casi di prognatismo mandibolare siano connessi ad un esagerato allungamento dell'abbozzo cartilagineo o ad un'alterata ossificazione della porzione dentale o angolare che concorrono a formare il corpo mandibolare. Nelle larve di branzino, la presenza delle strutture cartilaginee che compongono il cavo orale è visibile al termine del riassorbimento del sacco vitellino; l'inizio dell'ossificazione mandibolare avviene a 25 giorni dalla schiusa e risulta completa in tutti i suoi elementi ad 80 giorni (Kuzir *et al.*, 2004).

Problemi economici connessi al prognatismo

Il prognatismo mandibolare negli avannotti di branzino è compatibile con la vita, non deprime sensibilmente l'accrescimento durante le fasi di ingrasso, non aumenta la suscettibilità alle malattie. Sebbene il prognatismo tenda con l'accrescimento ad essere meno pronunciato rispetto alla fase di avannotto, viene ritenuta dagli allevatori una caratteristica estetica sgradevole per l'acquirente; considerato che il branzino viene proposto al dettaglio prevalentemente intero e non in filetti, la morfologia della testa rientra tra i principali parametri di valutazione per il cliente, determinando deprezzamento del prodotto. Gli avannotti prognati rappresentano per l'avannotteria un costo in termini di spazio d'allevamento, alimento somministrato, manodopera impiegata nel processo di selezione manuale, e causano un deprezzamento della partita in fase di vendita ai centri di ingrasso. La selezione e lo scarto dei soggetti malformati

rappresenta non solo un problema di natura economico-manageriale, ma anche un problema di benessere animale.

A livello comunitario, l'EFSA non si è ancora pronunciata in merito alla corretta forma di eutanasia da applicare ad avannotti moribondi o con deformità, tuttavia l'overdose di anestetico MS222 (Tricaina metansulfonato) sembra essere il metodo preferibile (EFSA, 2009). Certamente l'individuazione e risoluzione delle cause di questa malformazione è l'obiettivo da raggiungere.

Cause di malformazione di mandibola e mascella

Nell'uomo la prevalenza del prognatismo varia tra le popolazioni: è massima in Estremo Oriente (15%) e minima tra i caucasici (1%), e presenta una predisposizione familiare (Emrich *et al.*, 1965). Sebbene siano stati individuati diversi fattori ambientali (traumatismi, squilibri endocrini, difetti posturali ed infezioni) che contribuiscono al manifestarsi del difetto, si ipotizza che la causa principale possa essere di natura genetica, anche se il modello di ereditabilità per questo carattere è tuttora controverso. Probabilmente per la maggior parte delle famiglie affette valgono cause poligeniche e multifattoriali. Esistono tuttavia notevoli eccezioni tra le casate nobiliari europee; sugli Asburgo, un'analisi completa di *pedigree* ha evidenziato una trasmissione autosomica dominante a penetranza incompleta (Wolff *et al.*, 1993). Uno studio condotto su famiglie coreane e giapponesi ha associato tre loci situati su cromosomi differenti al manifestarsi della malformazione (Yamaguchi, 2005); tra i geni così evidenziati spiccano sequenze codificanti per fosfatasi alcalina, eparan-solfato proteoglicano II e matrilina-1 (proteina della matrice cartilaginea). Queste glicoproteine sono implicate nell'accrescimento del tessuto cartilagineo (eparan-solfato proteoglicano II), nella sua maturazione (matrilina-1) e mineralizzazione a tessuto osseo (fosfatasi alcalina).

Ad oggi, gli studi condotti sulle malformazioni nelle prime fasi di sviluppo in *D. labrax* non hanno evidenziato una causa univoca per il prognatismo mandibolare suggerendo di

fatto, per questo carattere, un' eziologia multifattoriale (Abdel *et al.*, 2004; Villeneuve *et al.*, 2005).

Sperimentazioni effettuate su larve di zebrafish (*Danio rerio*) hanno evidenziato un potenziale ruolo dei geni detti *Hedgehog Hh* (*Sonic Hedgehog a e b*) nello sviluppo normale della mandibola (Teraoka *et al.*, 2006). La via di segnale in cui questi geni sono coinvolti gioca un ruolo di rilievo nell'organogenesi dei vertebrati, e le proteine sintetizzate dai geni Hh agiscono come morfogeni. È stato dimostrato che mutazioni a carico di *Sonic Hedgehog* o trattamenti con composti tossici quali 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) riducono o alterano la via di segnale dei geni Hh, provocando l'alterazione del normale sviluppo mandibolare in senso ipoplastico. Ulteriori ipotesi chiamano in causa la via di segnale dei RAR (*Retinoic Acid Receptor*), recettori intracellulari dell'acido retinoico, come implicati nella malformazione e iposviluppo della mandibola. Trattamenti nelle fasi larvali di passera giapponese (*Paralichthys olivaceus*) con retinoidi sintetici, agonisti RAR, ha infatti indotto malformazioni in tutti i casi testati (Haga *et al.*, 2002).

Somministrazione di diete contenenti concentrazioni eccessive di vitamina A in avannotti di branzino ha determinato la comparsa di deformità del neurocranio e splancocranio; in particolare un *deficit* nello sviluppo di ossa premaxillari, maxillari, parasfenoidi, pterigoidei e del piatto etmoidale denominato "*pugheadness*" traducibile come "muso schiacciato". Questa deformazione è causata da iperproliferazione poiché determina un iposviluppo dell'arcata mascellare che rende l'arco mandibolare, normalmente sviluppato, falsamente sporgente (Mazurais *et al.*, 2009).

Studi condotti sul ruolo dei lipidi nello sviluppo larvale del branzino (Villeneuve *et al.*, 2005) hanno evidenziato la capacità degli acidi grassi di modulare la trascrizione di geni coinvolti nel loro metabolismo attraverso recettori nucleari PPAR (*Peroxisome Proliferator-Activated Receptor*); questi ultimi formano eterodimeri con altri recettori

nucleari, tra i quali gli RXR (Retinoid X Receptor) coinvolti nella via degli acidi retinoici (via del RAR). Quest'ultima controlla l'espressione di geni morfogenetici appartenenti a diverse famiglie: geni *Hox*, geni codificanti per BMP (*Bone Morphogenetic Protein*) e IGF (*Insuline-like Growth Factor*), ed interagisce con l'espressione dei geni *Hedgehog* (Suzuki *et al.*, 1998). Un incremento dell'attivazione dei recettori PPAR, indotto da un rapporto non ottimale tra fosfolipidi e lipidi neutri o un eccesso di PUFA Ω 3 nella dieta, determina una perturbazione della via RAR che risulta in un aumento nelle malformazioni rachidee e cefaliche e nella mortalità delle larve (Villeneuve *et al.*, 2005).

1.3.2 Principali patologie infettive negli allevamenti di orata e branzino

Le principali patologie infettive riscontrabili negli allevamenti di branzino e orata del Mediterraneo sono principalmente di natura batterica (Pasteurellosi, Flavobatteriosi, Vibriosi) e virale (Nodaviosi, Linfocisti). Parassitosi sono sostenute da ectoparassiti dei generi *Trichodina*, *Amyloodinium* e *Cryptocaryon*, dal trematode monogeneo *Diplectanum aequans* e da endoparassiti del genere *Sphaerospora*.

1.3.2.1 Patologie ad eziologia batterica

Le patologie batteriche "classiche", Vibriosi e Pasteurellosi, non hanno determinato negli ultimi anni ingenti problemi a livello nazionale; la seconda, tuttavia, è sempre presente nelle avannotterie e può determinare focolai con mortalità anche elevata.

La Pasteurellosi ittica, nota anche col nome di Fotobatteriosi (Gauthier *et al.*, 1995), è un'infezione sostenuta da *Photobacterium damsela* sottospecie *piscicida*, batterio Gram-negativo, alofilo, corto e robusto (0.5-1.3 x 0.7-4.0 μ m), non-mobile e asporigeno. La Pasteurellosi, comparsa in Italia nel 1991 (Ceschia *et al.*, 1991), determina gravi epizootie negli allevamenti di spigole ed orate soprattutto nelle stagioni calde, quando la temperatura dell'acqua raggiunge i 22-25°C, o quando la salinità si riduce al di sotto del 25-30‰ in seguito a prolungati periodi di pioggia. Condizioni

favorevoli ed aggravanti l'infezione sono l'elevata densità di popolazione, una bassa concentrazione di ossigeno disciolto nell'acqua, un elevato grado di inquinamento e la concomitante presenza di patogeni opportunisti. La via di trasmissione più probabile sembra essere il contatto diretto attraverso l'epitelio branchiale; è tuttavia considerato possibile anche il contagio per via orale attraverso l'alimento (Kawahara *et al.*, 1987). La Pasteurellosi ittica può essere considerata essenzialmente una patologia setticemica con manifestazioni sia acute che croniche. L'infezione a carattere acuto colpisce soprattutto soggetti molto giovani (tra 1 e 40 gr di peso corporeo) e determina in genere una mortalità molto elevata (80-90%). Nella forma acuta i pesci presentano nuoto atassico, anoressia e letargia, il corpo appare iperpigmentato e compaiono emorragie diffuse alla base delle pinne e sui fianchi. A livello viscerale, prevalentemente su fegato, milza e rene, si possono evidenziare lesioni necrotizzanti multifocali ed emorragie. L'epitelio branchiale presenta aree di necrosi e discrinia, causa di insufficienza respiratoria. L'infezione cronica colpisce più frequentemente soggetti adulti dove la mortalità scende fino a circa il 30%. In questa forma si possono osservare caratteristici noduli biancastri, di diametro normalmente compreso tra 0,5 e 3,5 mm, localizzati in reni e milza e, di dimensioni minori, anche a livello epatico (Hawke *et al.*, 1987).

Le Vibriosi in *D. labrax* sono sostenute prevalentemente da *Vibrio anguillarum*, meno frequentemente da altre specie come *V. ordalii*, *V. vulnificus* e *V. alginolyticus*.

Vibrio anguillarum è un batterio Gram negativo, normalmente presente nella microflora ambientale, di forma bastoncellare con tendenza ad incurvarsi lievemente, mobile per la presenza di un singolo flagello, aerobio-anaerobio facoltativo, alofilo. Focolai di Vibriosi si manifestano in concomitanza di eventi stressanti che riducono le difese immunitarie dei soggetti allevati, quali per esempio rapide variazioni giornaliere della temperatura nei mesi primaverili e autunnali, selezione manuale, sovraffollamento e

peggioramento della qualità dell'acqua. Sebbene tutte le classi d'età risultino colpite, gli avannotti durante la fase di pre-ingrasso risultano più suscettibili e presentano mortalità più elevata (35%) rispetto ad una media del 15% durante l'ingrasso. I pesci affetti da Vibriosi manifestano segni di setticemia emorragica con letargia, anoressia, melanosi, esoftalmo bilaterale, arrossamenti e spandimenti emorragici sottocutanei, evidenti soprattutto a livello del ventre, lieve anemia branchiale e discrinia. La vescica natatoria è frequentemente dilatata e porta vicino alla superficie i pesci moribondi. Letargia, inscurimento della livrea e stazione a pelo d'acqua sono i sintomi principali negli avannotti.

Le infezioni da Mixobatteri, note anche con il nome di Flavobatteriosi, sono malattie infettive emergenti in maricoltura, soprattutto nel bacino del Mediterraneo dove l'agente eziologico, *Flexibacter maritimus* (noto anche come *Tenacibaculum maritimum*), è stato isolato a partire dal 1993 in allevamenti di branzini, orate e dentici (Alsina *et al.*, 1993). Appartenente alla famiglia delle *Flavobacteriaceae*, è un patogeno opportunista, bastoncellare, Gram negativo, mobile per scivolamento (*gliding*), aerobio stretto, alofilo. La sintomatologia della Flavobatteriosi è caratterizzata dalla comparsa di erosioni cutanee accompagnate da depigmentazione ed emorragie, ulcere talora profonde sino ad interessare la muscolatura sottostante, ed erosioni delle pinne e della coda, cachessia e ritardo della crescita, talvolta con esito letale (Cepeda *et al.*, 2002).

1.3.2.2 Patologie ad eziologia virale

Tra le patologie ad eziologia virale che interessano il branzino si deve segnalare la presenza costante di focolai di Nodaviosi (Encefalopatia Retinopatia Virale), per quanto riguarda l'orata, invece, va segnalata l'insorgenza sporadica di Linfocisti.

La Nodaviosi è una malattia contagiosa sostenuta da un virus a RNA appartenente alla famiglia *Nodaviridae*, genere *Betanodavirus*. E' una patologia che interessa oltre 30 specie ittiche, sia di acqua marina che di acqua dolce, appartenenti agli ordini dei

Perciformi e Pleuronectiformi; ed è diffusa in tutti i continenti ad eccezione dell’Africa. In Europa è stata segnalata a partire dagli anni ‘90 in Francia, Grecia, Italia e Spagna. Gli episodi di malattia si manifestano in genere quando la temperatura dell’acqua raggiunge i 18-20°C e la sintomatologia si rende evidente al di sopra dei 22-25°C. La Nodaviriosi è caratterizzata da vacuolizzazione e necrosi a carico di retina e sistema nervoso centrale ed è associata ad un’elevata letalità (fino al 100%) negli stadi larvali; percentuale che va man mano diminuendo negli stadi giovanili al di sotto di un anno di età (20-30%), per assumere una forma a “stillicidio” negli individui adulti.

L’infezione viene generalmente introdotta in allevamento con l’inserimento di soggetti portatori asintomatici o per contatto con pesci selvatici infetti, trasmettendosi poi attraverso l’acqua contaminata, grazie all’elevata resistenza del virus; è dimostrata anche la trasmissione per via verticale della malattia. La malattia conclamata si può manifestare più facilmente, in una popolazione infetta, in concomitanza di fattori stressanti quali l’elevata densità di allevamento; tra i segni clinici si riscontrano: mortalità improvvisa senza sintomi nelle larve, atassia e nuoto circolare nei giovani, scatti improvvisi con la testa fuori dall’acqua. Nei soggetti adulti è meno evidente la sintomatologia nervosa, mentre più frequenti appaiono la letargia, cheratite e l’insorgenza di erosioni cutanee a livello del capo.

La Linfocisti è una patologia virale cronica e autolimitante che colpisce più di 125 specie di teleostei in tutto il mondo, tra cui l’orata. L’agente eziologico di questa patologia è un virus a DNA, appartenente al genere *Iridovirus*, noto col il nome di *Lymphocystivirus*, che replica nel citoplasma delle cellule affette. La trasmissione avviene generalmente per contatto con individui infetti; traumi a livello della cute dovuti a maneggiamento, eccessiva densità di popolazione o comportamenti aggressivi può favorire lo sviluppo e la rapida trasmissione dell’infezione.

Il *Lymphocystivirus* replica a livello dei fibroblasti della cute con la formazione di cellule iperplastiche che non vanno però incontro a mitosi; le lesioni, a carico della pelle o delle pinne, consistono in masse nodulari simili a verruche che possono andare dai 0,3 mm fino a superare i 2 mm di diametro. Raramente le lesioni coinvolgono organi interni. Trattandosi di una patologia autolimitante, queste lesioni tendono a guarire spontaneamente lasciando come unica conseguenza cicatrici a livello epiteliale. Sebbene questa patologia non determini un'elevata mortalità, da non sottovalutare è la probabilità di sviluppare infezioni secondarie di tipo batterico o micotico; le lesioni, inoltre, possono comunque determinare diverse conseguenze a seconda della loro localizzazione come, ad esempio, difficoltà respiratorie se localizzate nelle branchie o a cibarsi se nella zona periboccale. La presenza di cicatrici a livello della pelle, inoltre, va ad inficiare notevolmente la commerciabilità del prodotto determinando ingenti perdite economiche per l'allevamento colpito.

1.3.2.3 Patologie parassitarie

Le patologie parassitarie sono le patologie “emergenti”, che possono creare più problemi in particolari situazioni ambientali o casi sporadici. Quelle maggiormente segnalate sono tricotiniasi e oodiniasi; entrambe possono avere un forte impatto economico (soprattutto in vallicoltura); sono quasi assenti negli allevamenti con ricircolo.

I mixosporidi (in particolare *Myxidium leei*, che appare quello dotato di maggiore patogenicità) stanno assumendo tra le malattie parassitarie dell'orata un ruolo di primaria importanza.

Altre parassitosi (infestioni da trematodi monogenesi) possono destare preoccupazione in allevamenti in cui non vengono operate adeguate forme di prevenzione o soggetti a particolari condizioni ambientali.

1.4 La genomica funzionale al servizio dell'acquacoltura

1.4.1 Cos'è la genomica funzionale

Tutti gli esseri viventi portano nel loro genoma istruzioni ben precise sulla loro struttura, crescita e funzione. La genomica è il settore delle scienze biologiche che si occupa di studiare e decodificare queste istruzioni.

La nascita della genomica è recente, solitamente viene ricondotta al sequenziamento del primo genoma completo, le 5.375 bp del fago PHI-X174, avvenuto nel 1977 (Sanger *et al.* 1977) . Da quel momento la genomica ha subito uno sviluppo esponenziale contando ai giorni nostri oltre 2.000 genomi procarioti e poco più di 200 genomi eucarioti completi o quasi completi, e altrettanti in fase di sequenziamento (<http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html>).

Sebbene la funzione del DNA sia quella di custodire l'informazione ed assicurarne la corretta trasmissione alle generazioni successive, questa viene costantemente modulata da processi dinamici che ne influenzano l'attività e l'effetto. La difficoltà maggiore sta nel comprendere le complesse interazioni tra i geni, i loro prodotti e l'ambiente.

La genomica funzionale è una branca delle scienze che ha lo scopo di determinare quale gene è responsabile di un determinato cambiamento/carattere fenotipico, permettendo di capire i meccanismi di controllo molecolare dei processi biologici e fisiologici, della risposta alle patologie e agli stimoli ambientali. La genomica funzionale ha proprio come obiettivo quello di comprendere la funzione di ciascun gene, come esso viene regolato e come esso interagisce con altri geni. In opposizione all'approccio "statico" della genomica, la genomica funzionale si focalizza in aspetti dinamici come la trascrizione e la traduzione genica nonché le interazioni tra proteine.

L'applicazione della genomica, della trascrittomica, della proteomica e di nuove "omics" emergenti può offrire un contributo tangibile al miglioramento del benessere animale nonché alla comprensione dei processi biologici che determinano le

caratteristiche fenotipiche (tasso di crescita, presenza/assenza deformità, suscettibilità patologie ecc.) che hanno un'importanza considerevole in acquacoltura, sia dal punto di vista economico che qualitativo.

1.4.2 Contributo della genomica funzionale all'acquacoltura

Come già anticipato in precedenza, l'acquacoltura e la maricoltura sono considerate essere i principali settori destinati a guidare l'incremento delle produzioni animali in Europa e nel mondo. Molte di queste attività, tuttavia, hanno ancora oggi una gestione di tipo tradizionale o comunque scarsamente tecnologica. Nonostante sia l'allevamento di branzino che quello di orata abbiano registrato un notevole sviluppo a partire dagli anni '80, diverse sono le limitazioni che ancora permangono. Tra queste rientrano l'elevata mortalità larvale e la considerevole frequenza di malformazioni scheletriche; queste anomalie di sviluppo compromettono notevolmente la produzione, del branzino in particolare, determinandone una ridotta *fitness* e/o commercializzazione con conseguenze economiche rilevanti. Un'altra delle problematiche principali che coinvolgono l'allevamento ittico è l'insorgenza di patologie. Negli allevamenti intensivi in particolar modo, i pesci allevati sono suscettibili ad un elevato numero di infezioni batteriche, virali e parassitarie che determinano ingenti perdite e un grosso impatto nella produzione sia del il branzino che dell'orata.

Una miglior conoscenza dei meccanismi molecolari che sottendono a caratteristiche produttive (tasso di crescita, sviluppo muscolare e osseo, suscettibilità a stress o patologie ecc.) offre la prospettiva di rivoluzionare le produzioni animali gettando le basi per migliorati programmi di selezione nonché per un effettivo monitoraggio degli effetti che le condizioni di allevamento hanno sugli animali. La genomica funzionale, ed in particolare gli approcci "*whole-genome*" per lo studio delle interazioni tra i geni e l'ambiente circostante, offre nuove opportunità per raggiungere questo obiettivo; non a

caso numerosi progetti di ricerca in questo ambito hanno recentemente preso piede per molte specie di allevamento.

Negli ultimi hanno avuto un forte sviluppo studi di genetica e genomica su specie ittiche modello quali zebrafish (*Danio rerio*, Teh *et al.* 2005), tetraodon (*Tetraodon nigroviridis*,; Crollius, 2006) e fugu (*Takifugu rubripes*; Clark *et al.* 2001). Recentemente i genomi di altre due specie di teleostei; stickleback (*Gasterosteus aculeatus*) e medaka (*Oryzias latipes*, Takeda *et al.* 2008) sono stati interamente sequenziati (<http://www.ensembl.org>). Il rapido sviluppo della genomica funzionale nelle specie ittiche considerate modello apre la strada a validi studi di tipo comparativo; con l'accumulo di informazioni su queste specie, infatti, la caratterizzazione anche di specie non modello, ma comunque di forte interesse economico, è ora possibile.

Il branzino e l'orata sono due specie di grande interesse commerciale, in particolar modo per l'acquacoltura del Mediterraneo. Entrambi appartengono all'ordine dei Perciformi, il più ricco in numero di specie (circa 10.000), tra cui numerose impiegate nell'acquacoltura, ma piuttosto povero in termini di informazioni e strumenti genomici al momento disponibili.

Proprio il ruolo di grande importanza che *D. labrax* e *S. aurata* ricoprono nell'acquacoltura marina li rende potenziali specie modello a loro volta e ha motivato un crescente numero di studi scientifici relativi a vari ambiti tra i quali l'immunologia, l'endocrinologia, la morfologia ossea, la fisiologia muscolare e la genetica molecolare.

L'interesse crescente è documentato dal considerevole aumento del numero di pubblicazioni ad essi relative. Il corredo di strumenti genomici disponibili per *Sparus aurata* è in costante evoluzione portando allo sviluppo nel 2005 di un microarray a cDNA di prima generazione per lo studio dell'espressione genica (Sarropoulou *et al.*, 2005). Grossi vantaggi verrebbero tratti, inoltre, dall'identificazione di marcatori genetici, correlabili a tratti di interesse come il tasso di crescita e la resistenza alle

patologie, che permettano di effettuare una “selezione assistita da marcatori” o MAS (*Marker Assisted Selection*). Ingenti sforzi per raggiungere questo scopo hanno portato alla definizione di una mappa di ibridi di radiazione (RH, *Radiation Hybrid*) preliminare (Senger *et al.*, 2006), sviluppata e portata a più di 1000 marcatori, ed una mappa RH di seconda generazione (Sarroupolou *et al.*, 2007); nonché allo sviluppo di una mappa di *linkage* di prima e di seconda generazione (Franch *et al.* 2006).

Oltre a studi destinati alla “*breeding selection*”, grossa attenzione è stata rivolta alla caratterizzazione di geni coinvolti nella risposta immunitaria nonché al sequenziamento su media scala di EST (*Expressed Sequence Tag*) che, in organismi per i quali le informazioni sul genoma sono ancora limitate, forniscono una ricca risorsa per l’identificazione e la caratterizzazione di nuovi geni. Nel corso di questo progetto di dottorato è stato inoltre sviluppato e validato una piattaforma microarray a oligonucleotidi (Ferraresso *et al.* 2008) sulla base di poco più di 19.000 trascritti unici di orata.

Anche per il branzino il numero di strumenti e di informazioni disponibili è in continuo aumento. Come per l’orata, un grosso interesse è rivolto all’identificazione di marcatori molecolari di caratteristiche vantaggiose dal punto di vista economico; questo ha permesso recentemente la costruzione di una mappa di *linkage* anche per *D. labrax* (Chistiakov *et al.* 2005, 2008). Sforzi ingenti sono tuttavia diretti al sequenziamento del genoma di questa specie, un prima libreria BAC (*Bacterial Artificial Chromosome*) è stata costruita già nel 2006 (Whitaker *et al.* 2006) ed è ora in corso il sequenziamento completo dell’intero genoma di branzino (Kuhl H., comunicazione personale).

Ad oggi, la conoscenza dei geni del sistema immunitario responsabili della resistenza di questa specie alla patologie è piuttosto povera ma sta avendo un rapido sviluppo. Negli ultimi anni, infatti, un elevato numero di studi si sono incentrati nella caratterizzazione del sistema immunitario del branzino (Chistiakov *et al.* 2007) e della sua risposta ad

infezioni sia virali (Scapigliati *et al.* 2009) che batteriche (Sarropoulou *et al.* 2009, Sepulchre *et al.* 2007, Nascimento *et al.* 2007).

Nel corso di questo progetto di dottorato gli sforzi per aumentare la conoscenza e gli strumenti genomici a disposizione per *D. labrax* sono stati diretti su due fronti: un primo studio si è incentrato sull'identificazione e la caratterizzazione di una famiglia di recettori appartenenti al sistema immunitario innato del branzino, i *Novel Immune-type Receptor* (NITR) (Ferraresso *et al.* 2009) mentre un secondo progetto si è dedicato alla messa a punto di un microarray a oligo-DNA per lo studio dell'espressione genica in questa specie, alla sua validazione e alla sua applicazione nell'analisi dei profili di espressione genica in esemplari affetti da prognatismo (Ferraresso *et al.*, articolo sottomesso).

1.4.2.1 Approcci high-throughput: DNA microarray

Lo studio dei profili di espressione genica mediante la tecnologia *microarray* è tra le tecnologie maggiormente utilizzate nella branca della genomica funzionale. I *microarray* a DNA possono essere descritti come una serie ordinata di frammenti di acidi nucleici noti posizionati su un supporto solido e che fungono da “*detector*” molecolari. L'ibridazione di molecole di DNA o RNA fluorescenti su queste *slide*, contenenti migliaia “*detector*”, permette di identificare e quantificare un numero elevatissimo di molecole di RNA o DNA alla volta.

A partire dalla loro prima descrizione ormai 15 anni fa (Schena *et al.* 1995), i *microarray* a DNA sono diventati degli strumenti di routine in molti laboratori di tutto il mondo. Ad oggi, i *microarray* vengono utilizzati per diversi scopi tra cui *Comparative Genomics Hybridization* (CGH) (Oostlander *et al.* 2004), ChIP-on-CHIP (Nègre *et al.* 2006), *genotyping* (Hacia 1999) e quantificazione di microRNA (Yin *et al.* 2008); tuttavia la loro principale applicazione rimane l'analisi dell'espressione genica su larga scala.

A seconda delle dimensioni e della loro complessità, le piattaforme *microarray* sono strumenti utili per identificare la risposta dell'intero trascrittoma a diverse condizioni fisiologiche o sperimentali. I *microarray* per lo studio dei profili di espressione possono essere fabbricati mediante lo spottaggio di cloni di cDNA o prodotti di PCR (*Polymerase Chain Reaction*) o mediante la sintesi, direttamente sul supporto, di oligonucleotidi.

Sia le piattaforme a cDNA che quelle a oligo-DNA permettono l'analisi di migliaia di geni contemporaneamente, rendendo possibile una descrizione globale del fenomeno sotto osservazione e permettendo, quindi, di definire le alterazioni geniche che stanno alla base di condizioni patologiche o di identificare i profili di espressione che caratterizzano diverse condizioni di interesse. Assieme ad altri strumenti di genomica funzionale *high-throughput*, lo studio dell'espressione genica tramite *microarray* permette un approccio obiettivo ed imparziale (libero da ipotesi iniziali) a diversi tipi di problematiche, dove l'unico assunto di base è che i geni rappresentati nella piattaforma cambino il livello di espressione in relazione alle diverse condizioni sperimentali.

Nell'era del sequenziamento come può essere considerata quella attuale, numerose piattaforme *microarray* sono state sviluppate, oltre che per l'uomo, per un elevato numero di specie "modello". Poiché fino a poco tempo fa non venivano considerate modelli sperimentali "ideali", questo genere di tecnologia non era stata estesa alle specie ittiche a eccezione di una ridotta elite rappresentata da *D. rerio*, *T. nigroviridis*, *T. rubripes* e *O. latipes*.

In breve tempo questa tecnologia ha, tuttavia, avuto uno sviluppo esponenziale portando alla progettazione e applicazione di piattaforme *microarray*, sia a cDNA che a oligonucleotidi, per quasi 20 diverse specie di teleostei (Miller *et al.* 2008) appartenenti a 3 diversi ordini (*Ostariophysi*, *Protacanthopterygi*, *Acanthopterygi*).

1.4.2.2 Studio mirato di geni candidati: il sistema immunitario dei teleostei

Alla luce della sempre crescente importanza che il settore dell'acquacoltura sta ricoprendo in Europa e nel mondo negli ultimi decenni, un particolare interesse è stato rivolto di recente al sistema immunitario dei teleostei. Lo scopo è quello di far luce sui ruoli che esso ricopre nella risposta delle specie allevate alle patologie e alle condizioni di stress, che caratterizzano in particolar modo gli allevamenti a carattere intensivo.

Poiché i teleostei sono ancora poco conosciuti dal punto di vista immunitario, diversi approcci di genomica funzionale possono fornire le informazioni molecolari necessarie sul sistema immunitario di queste specie.

La risposta immunitaria innata è un sistema di difesa primitivo esistente sia nelle piante che negli animali. Nei vertebrati è un sistema complesso composto di una duplice componente; cellulare e umorale. I teleostei, la cui separazione dai tetrapodi risale a circa 450 milioni di anni fa, possiedono alcune componenti dell'immunità innata che mostrano un grado di conservazione considerevole con i vertebrati superiori. Tuttavia, c'è un numero sempre maggiore di evidenze che illustrano come i teleostei possiedano componenti e funzioni non osservate nei mammiferi (Aoki *et al.* 2008) suggerendo, quindi, un certo grado di diversità e di complessità nella funzione del sistema immunitario innato dei pesci. La conoscenza delle relazioni e delle funzioni che queste molecole immuno-correlate hanno in specie ittiche modello e/o di interesse economico è quindi importante, e i nuovi strumenti forniti dalla genomica possono permettere l'ottenimento di queste informazioni.

Diversi approcci sono ad oggi utilizzati per l'identificazione e la caratterizzazione di geni appartenenti al sistema immunitario; quello principalmente utilizzato in un elevato numero di specie è l'analisi di ESTs. Le EST sono sequenze di cDNA ottenute da librerie che spesso rappresentano un singolo tessuto o un organismo soggetto ad una particolare condizione sperimentale. Come tali le EST provvedono un'istantanea dello

stato fisiologico di un organismo; inoltre, per le specie per le quali le informazioni genomiche sono limitate, queste sequenze sono una preziosa risorsa per l'identificazione di nuovi geni. L'analisi di EST può quindi permettere l'identificazione di geni coinvolti nella risposta immunitaria sulla base di geni noti e strutturalmente correlati in altre specie.

Anche tecnologie *high-throughput*, impiegate come strumenti per la caratterizzazione di geni e prodotti di geni, sono diventati una strategia importante per fare chiarezza sull'immunologia dei teleostei; in particolar modo se questi approcci permettono il rapido accumulo di dati in assenza di una precedenti nozioni sulla risposta immunitaria (Robalino *et al.* 2009). Trascritti che mostrano consistenti alterazioni nei livelli di espressione in seguito ad infezioni sperimentali, ad esempio, possono essere considerati geni candidati per successivi studi funzionali.

Altre risorse che possono essere impiegate per l'identificazione di componenti immuno-correlate; sono le librerie di BAC (*Bacterial Artificial Chromosome*) e le mappe genomiche. Analogamente alle EST, le sequenze BAC forniscono il materiale grezzo per la ricerca comparativa di geni candidati o per lo studio di particolari loci immunitari (per es. peptidi antimicrobici, loci MHC, NITR spesso organizzati in *tandem* nel genoma).

L'analisi approfondita delle componenti dell'immunità innata dei teleostei, comprensiva di relazioni e funzione, offre certamente la prospettiva di generare nuove tecnologie applicabili al miglioramento dell'acquacoltura.

1.5 Scopo del lavoro

L'acquacoltura ha raggiunto quest'anno il traguardo del 50 per cento della produzione di tutto il pesce consumato nel mondo ed è considerata essere il settore destinato a guidare l'incremento delle produzioni animali in Europa e nel mondo. Sebbene questa industria abbia avuto uno sviluppo enorme rispetto al passato, diverse limitazioni non sono ancora state superate; in particolar modo l'elevata mortalità larvale, la forte suscettibilità a stress e patologie nonché lo sviluppo di malformazioni scheletriche. La genomica funzionale, incrementando la conoscenza dei meccanismi molecolari responsabili di caratteristiche fenotipiche di forte interesse economico e qualitativo in acquacoltura, offre la prospettiva di rivoluzionare e rendere più efficiente questo settore. L'obiettivo di questo triennio di dottorato è stato quello di sviluppare nuovi strumenti genomici ed applicarli allo studio di due specie di grande importanza nell'acquacoltura europea, *S. aurata* e *D. labrax*. Lo scopo è stata la caratterizzazione di componenti e meccanismi che giocano un ruolo in tratti importanti come appunto lo sviluppo di malformazioni scheletriche e la suscettibilità alle patologie.

In tal senso, gli sforzi sono stati diretti su tre fronti:

- Il sistema immunitario ricopre un ruolo chiave nella risposta delle specie allevate alle patologie; c'è quindi un grande impegno nel migliorare la conoscenza delle caratteristiche e delle funzioni delle molecole immuno-correlate. In questo frangente, metodiche di biologia molecolare sono state utilizzate per l'identificazione e la caratterizzazione di una famiglia di recettori appartenenti al sistema immunitario del branzino, i *Novel Immune-type Receptor* (NITR). In totale sono stati identificati 27 geni e 3 pseudogeni, organizzati in *cluster* in una regione genomica di circa 270 kb; i membri di questa *cluster* sono caratterizzati da una elevata variabilità sia strutturale che a livello di organizzazione genomica che ha determinato la loro distinzione in 18 famiglie. Analisi comparative dei

NITR di branzino con gli omologhi identificati nelle altre specie di teleostei sono state inoltre condotte allo scopo di ottenere nuove informazioni riguardo l'organizzazione, la funzione e l'evoluzione di questi recettori dell'immunità innata.

- Nel corso di questo progetto di dottorato, quasi 20.000 trascritti unici di orata, derivanti da un progetto di sequenziamento su media scala di EST (*Expressed Sequenced Tag*), sono stati utilizzati per lo sviluppo di un *microarray* a oligonucleotidi per lo studio dei profili di espressione genica in *S. aurata*. Il *design* della piattaforma ha previsto la sintesi *in situ* di due sonde (60meri) indipendenti per ciascun trascritto utilizzando la tecnologia Agilent *SurePrint*TM. Sia l'accuratezza che la riproducibilità della piattaforma *microarray* sono state valutate mediante lo studio dei profili di espressione genica di due stadi precoci di sviluppo di orata, rispettivamente *pool* di larve raccolte a uno e quattro giorni dalla schiusa.
- Una seconda piattaforma *microarray* per la quantificazione dell'espressione genica è stata progettata per *D. labrax*. Analogamente a quanto già fatto per l'orata, anche per branzino una collezione di poco più di 19.000 trascritti unici sono stati impiegati per la costruzione di un *microarray* a oligonucleotidi sempre sfruttando la tecnologia Agilent *SurePrint*TM. Il *microarray* è stato quindi impiegato per analizzare i profili di espressione genica in mandibole e teste di esemplari di branzino affetti da prognatismo, una malformazione scheletrica che ne compromette pesantemente la produzione e la commerciabilità. Analisi statistiche dei risultati sperimentali sono state condotte allo scopo di identificare i trascritti la cui espressione risulti essere significativamente alterata negli individui colpiti da questa deformità.



**2. IDENTIFICAZIONE E CARATTERIZZAZIONE
DEI *NOVEL-IMMUNE TYPE RECEPTOR*
NEL BRANZINO**

2.1 INTRODUZIONE

Una sempre crescente attenzione è rivolta al ruolo che i processi innati hanno all'interno del sistema immunitario, alla varietà e diversità dei recettori dell'immunità innata, la loro interazione con le molecole di superficie e i meccanismi impiegati per la trasduzione dei segnali intracellulari.

Nei mammiferi sono stati identificati diversi *cluster* di recettori appartenenti al sistema immunitario innato; nei primati, particolare attenzione viene ora rivolta a recettori appartenenti alla superfamiglia delle immunoglobuline (IgSF; *Immunoglobulin Superfamily*): i *Killer Ig-like receptor* (KIR) (Martinez-Borra *et al.* 2008), gli *Ig-like transcript* (ILT) (Samadiris *et al.* 1997), i *Leukocyte-associated Inhibitory Receptors* (LAIR) (Meyaard *et al.* 1997) e gli *Fc γ Receptors* (Fc γ R). Tutti questi geni, strutturalmente e funzionalmente correlati, sono codificati in una regione genomica circoscritta del cromosoma umano 19q13.4, chiamato *Leukocyte Receptor Complex* (LRC) (Torkar *et al.* 1998), e anche nei primati non umani risultano essere organizzati in *tandem* (Canavez *et al.* 2001).

I recettori KIR, membri del *Natural Killer Complex* (NKC), sono stati ipotizzati interagire con le molecole MHC di classe I, permettendo in questo modo alle *Natural Killer* (NK) di distinguere le cellule bersaglio che espongono il “*non-self*” (Yokoyama *et al.* 2003). Questi recettori sono proteine transmembrana di tipo I caratterizzate dalla presenza di un numero variabile di domini immunoglobulinici (Ig) extra-cellulari del tipo C2 (V/C2 o I) o del tipo V (V). Questi domini Ig sono seguiti da un peptide di connessione (*stem region*), una regione transmembrana (TM) e una coda citoplasmatica (Cyt). A seconda delle caratteristiche di quest'ultima, i KIR possono essere suddivisi in recettori di tipo inibitorio o attivatore.

Nonostante l'eterogeneità strutturale dei vari recettori che interagiscono con le molecole MHC di classe I, il meccanismo responsabile della loro attività inibitoria è infatti

comune. Nella coda intracitoplasmatica di un recettore inibitorio sono presenti una o due sequenze contenenti residui di tirosina, denominate ITIM (*Immunoreceptor Tyrosine-based Inhibitory Motifs*), che in seguito a interazione con il ligando MHC I vengono fosforilate. Il dominio ITIM fosforilato è quindi responsabile del reclutamento di tirosino-fosfatasi, e in particolare di SHP-1, responsabile della propagazione del segnale negativo (Long, 2008).

Una caratteristica comune di tutte le famiglie di recettori MHC I è quella di contenere molecole che presentano differenze strutturali non solo a livello dei domini extracellulari, ma anche a livello delle regioni intracitoplasmatiche. Esistono infatti recettori della superfamiglia Ig che mancano del dominio ITIM nella coda intracitoplasmatica e presentano aminoacidi carichi positivamente nella regione transmembrana. Queste isoforme non inibitorie si associano a proteine adattatrici, quali DAP12, costituite da catene che contengono motivi di tipo ITAM (*Immunoreceptor Tyrosine-based Activation Motif*) e che, in seguito a interazione con le molecole MHC di classe I, sono responsabili dell'attivazione della citotossicità naturale e della produzione di citochine da parte delle cellule NK.

Nell'ultima decade, due famiglie geniche codificanti recettori *Ig-like*, non soggetti a riarrangiamento e con struttura molecolare simile a quella dei KIR sono state identificate in diverse specie di teleostei: i *Novel Immune-type Receptor* (NITR) e i *Novel Ig-like Transcript* (NILT).

Non molte sono, al momento, le informazioni a disposizione sui recettori NILT; finora questi geni sono stati identificati in sole due specie: due geni NILT sono stati isolati nella carpa comune (*Cyprinus carpio*) (Stet *et al.* 2005) e quattro nella trota (*Oncorhynchus mykiss*) (Kock *et al.* 2008; Østergaard *et al.* 2008). Questi recettori sono caratterizzati dalla presenza di uno o due domini Ig di tipo V seguiti da un dominio

transmembrana e da una coda citoplasmatica contenente un motivo ITAM o un numero variabile di domini ITIM.

Maggiori sono, invece, le informazioni a disposizione sui geni NITR. Questi recettori sono stati identificati per la prima volta nel 1999 in *Sphaeroides nephelus* (Strong *et al.* 1999), nel quale sono stati isolati 26 diversi geni tutti posizionati in un'unica regione genomica di circa 130 kb; a questa è seguita l'identificazione e caratterizzazione di quattro ulteriori geni nella trota (Yoder *et al.*, 2002), 18 nel pesce gatto (Hawke *et al.* 2001; Evenhuis *et al.* 2007), e un unico trascritto nell'halibut *Paralichthys olivaceus* (Piyaviriyakul *et al.* 2007).

Questi recettori sono stati ampiamente studiati e caratterizzati in zebrafish e medaka; in *D. rerio*, XX geni, raggruppati in 12 diverse famiglie sulla base del grado di identità aminoacidica ($\geq 70\%$) del dominio V, sono stati identificati nel cromosoma (chr.) 14, e due ulteriori famiglie nel chr. 7 (Yoder *et al.* 2001 and 2004), mentre, in *O. latipes* sono state identificate 24 famiglie recettoriali organizzate in 3 cluster localizzati rispettivamente sui cromosomi 10, 18 e 21 (Desai *et al.* 2008).

Trascritti codificanti recettori NITR sono stati identificati anche in diversi salmonidi, tra cui salmone (*Salmo salar*), merluzzo (*Gadus morhua*), halibut (*Atlantic halibut*) e coregone (*Coregonus clupeaformis*), nonché in stickleback (*Gasterosteus aculeatus*) (Yoder, 2009) noto in Italia col nome di spinarello.

La maggior parte dei geni NITR caratterizzati finora possiedono un dominio Ig di tipo variabile (V) e uno di tipo I nel comparto extracellulare. Entrambi questi domini possono contenere un motivo *Joining* (J) o *J-like*; tutte le evidenze raccolte finora dimostrano che queste sequenze (V-J o I-J) sono codificate da un unico esone, escludendone quindi la possibilità di ricombinazione.

La presenza del dominio extracellulare V è l'unica caratteristica comune a tra tutti i recettori NITR identificati sino ad oggi. Questo dominio risulta fortemente correlato ai

domini V delle Immunoglobuline e dei *T-cell receptors* (TCR) tramite residui conservati; aminoacidi in posizioni canoniche, infatti, risultano essere condivisi coi geni appartenenti alla IgSF. I domini V dei NITR presentano anche 3 regioni ipervariabili corrispondenti alle *Complementarity Determining Region* (CDR) presenti sia nelle Ig che nei TCR e deputate al riconoscimento del ligando (Rudolph *et al.* 2002).

Poiché un numero elevato di famiglie recettoriali possiedono uno o più domini V, la sola presenza di questo dominio non è sufficiente a definire un recettore come NITR; la caratteristica che permette la loro classificazione è la contemporanea presenza, nel comparto extracellulare, di un secondo dominio Ig di tipo I. I domini I sono costituiti da 6 cisteine conservate e anch'essi presentano alcuni dei residui conservati tra i domini V dei NITR, i TCR e le Ig, inclusi i due residui di cisteina (C₂₃ e C₁₀₄) ipotizzati essere responsabili del ponte disolfuro che stabilizza il *fold*ing caratteristico delle Ig.

I NITR costituiscono un complesso genico numeroso e altamente diversificato; strutturalmente, infatti, questi recettori possono essere distinti in base a:

- Numero dei domini Ig extracellulari (del tipo V e del tipo I);
- Numero di esoni codificanti i domini extracellulari;
- Presenza e localizzazione di motivi *joining* (J) o *J-like*;
- Presenza e numero dei domini ITIM nella coda citoplasmatica;
- Presenza di residui carichi positivamente nella regione transmembrana.

I geni NITR sono stati ipotizzati essere gli omologhi funzionali dei KIR mammiferi in quanto entrambi sembrano dare il via alla stessa cascata di segnale (Yoder 2009).

Analogamente agli *NK receptor* mammiferi, infatti, i NITR possono essere classificati in due gruppi funzionali:

- NITR inibitori, presentanti domini ITIM o *ITIM-like* nella coda citoplasmatica;
- NITR attivatori, caratterizzati dalla presenza nella regione transmembrana di residui carichi positivamente.

E' stato infatti osservato in zebrafish che i NITR attivatori sono in grado di dirigere la citolisi cellulo-mediata delle NK mediante interazione con l'ortologo teleosteo del gene DAP12 (Yoder *et al.* 2004), mentre un'azione inibitoria nei confronti delle NK è stata ipotizzata per i recettori che possiedono i domini ITIM e ITIM-like. Questa azione verrebbe esercitata mediante il reclutamento delle tirosina fosfatasi SHP-1 e SHP-2 e/o della inositolo fosfatasi SHIP-1 che darebbero il via alla cascata di segnali che include l'inibizione della MAPK e conseguentemente della funzione stessa delle NK (Yoder *et al.* 2007; Piyaviriyakul *et al.* 2007).

La stretta analogia tra i NITR e i membri della IgSF trova ulteriore conferma dall'analisi della struttura cristallografica, recentemente risolta, dei NITR; il sito di interazione con il ligando risiede infatti nella regione analoga alla CDR1 (*Complementarity Determining Region 1*) del dominio extracellulare V, mutazioni in questa sede dimostrano inoltre la sua potenziale interazione con un set polimorfico di ligandi (Cannon *et al.* 2008), analogamente a quanto avviene per i membri del *Natural Killer Complex*.

I ligandi dei *Novel Immune-type receptor* non sono ancora stati descritti e la natura fortemente polimorfica di molte specie di teleostei potrebbe rendere ulteriormente complicata la loro identificazione. La similarità strutturale tra i domini V di NITR e TCR, tuttavia, suggerisce che la natura di questi ligandi possa essere analoga; inoltre, l'elevato polimorfismo che contraddistingue le regioni ipervariabili dei domini V, fa supporre che i NITR stessi, come famiglia recettoriale, riconosca un ampio spettro di ligandi. E' stato proposto che, analogamente a quanto accade per alcuni dei recettori mammiferi, i NITR inibitori possano riconoscere le molecole MHC come segnale del "self", mentre i NITR attivatori potrebbero riconoscere le molecole MHC alterate e/o proteine virali o stress-correlate.

Nonostante la somiglianza strutturale dei NITR con alcuni membri del NKC, la loro origine è tuttora poco chiara. Analisi genomiche comparative tra i *cluster* genici dei NITR ed il *Leukocyte Receptor Complex* dei mammiferi non hanno mostrato che una debole sintenia, non è quindi da escludere che la presenza dei NITR possa essere ristretta ai soli teleostei (Yoder 2009). Le informazioni raccolte finora dimostrano che i NITR costituiscono una famiglia genica estremamente differenziata ed in rapida evoluzione. Diverse strutture genomiche come anche un diverso contenuto genico in termini di numerosità, sono stati osservati tra le diverse specie di teleostei, suggerendo che duplicazioni specie-specifiche e espansioni di famiglie geniche siano alla base dell'evoluzione del complesso genico dei NITR (Litman *et al.* 2001; Yoder 2009).

Il valore commerciale e l'importanza che il branzino ricopre nell'acquacoltura europea hanno determinato un'incremento esponenziale degli studi su questa specie in diverse aree di ricerca (l'immunologia, la fisiologia del muscolo, l'endocrinologia ecc.). Il sistema immunitario di *D. labrax* è uno dei meglio caratterizzati tra i teleostei e i Perciformi in particolare (Scapigliati *et al.* 2002; Randelli *et al.* 2009), il che lo renderebbe un buon candidato come specie modello per studi di tipo immunologico, complementariamente alle specie tradizionalmente considerate modello per questo scopo, che sono zebrafish e il pesce gatto, appartenenti all'ordine degli *Ostariophysi*.

In questo studio viene descritta l'identificazione e la caratterizzazione di 27 recettori NITR nel branzino (*D. labrax*). Tutti i 27 geni sono codificati in un'unica regione genomica di circa 270 kb e sono stati classificati in 18 diverse famiglie, confermando anche in questa specie l'elevata variabilità che caratterizza questo complesso multi genico. La risoluzione di questo complesso genico, assieme alle precedenti caratterizzazioni più approfondite in zebrafish, medaka e *Sphoeroides nephelus*, hanno reso possibile un'analisi evolutiva piuttosto esaustiva e la ricerca di una sintenia mantenuta tra le diverse specie.

2.2 MATERIALI E METODI

2.2.1 Campioni biologici ed estrazione dell'RNA

Quattro esemplari giovanili (17-20 gr) di *D. labrax* sono stati forniti dall'acquario dell'Istituto Zooprofilattico Sperimentale delle Venezie (Padova) e sacrificati con un eccesso di anestetico. Un totale di otto tessuti/organi (branchie, milza, intestino, muscolo scheletrico, pelle, rene cefalico, fegato e sangue) sono stati raccolti da ciascun animale e conservati in *RNALater*® (Ambion, Austin, TX), inizialmente a 4°C per 24 ore e successivamente a -20°C per lo stoccaggio a lungo termine.

L'RNA totale è stato estratto, a partire da poco più di 15 gr di tessuto, utilizzando il kit *RNAeasy Mini Kit* (Qiagen, Hilden, Germany) seguendo le indicazioni del produttore. La qualità di ciascun estratto è stata poi valutata mediante elettroforesi su gel di agarosio (1%) contenente *SYBR Safe™ DNA Gel stain 10,000X* (Invitrogen™, Carlsbad, California).

2.2.2 Sequenziamento dei cloni BAC di *D. labrax*

Il sequenziamento dei BAC (*Bacterial Artificial Chromosome*) contenenti il *cluster* genico dei NITR è stato realizzato presso il Max Planck Institute for Molecular Genetics di Berlino, di seguito ne vengono riportate brevemente le diverse fasi. Un primo BAC candidato (bassbac-18k1), potenzialmente contenente geni NITR, è stato identificato mediante analisi di Blast (*Basic Local Alignment Search Tool*) su una serie di *BAC-ends* sequenziate da una libreria di BAC di *D. labrax* (Kuhl *et al. in preparazione*). In seguito al sequenziamento *shot-gun* di bassbac-18k1, un secondo BAC (bassbac-79e10), contenente un frammento contiguo della regione genomica candidata in base ad analisi di Blast, è stato sequenziato per intero.

Il DNA di questi cloni BAC è stato isolato per mezzo di lisi alcalina e successivamente il DNA di *E. coli* rimanente è stato rimosso mediante digestione con esonucleasi ATP-dipendente. Il DNA purificato è stato quindi frammentato tramite ultrasuoni e molecole

delle dimensioni di 1-4 kb sono state selezionate per l'*end-polishing* con la T4-DNA-polimerasi e la DNA-polimerasi I (Klenow). I frammenti ottenuti sono stati quindi ligati con l'utilizzo della T4-DNA ligasi all'interno del sito di restrizione *SmaI* del vettore di sequenziamento puC19. Le cellule competenti di *E. coli* DH10B sono state quindi trasformate tramite elettroporazione. Per ciascun BAC è stata costruita una libreria con un *coverage* di circa 10 volte e il DNA plasmidico è stato purificato per il sequenziamento con ABI BigDye v3.1 Terminator (Applied Biosystem™, Foster City, California).

Le sequenze grezze sono state processate mediante l'utilizzo del programma PHRED (Ewing *et al.* 1998) e la rimozione della sequenza del vettore, così come delle sequenze di bassa qualità, è stata eseguita con LUCY (Chou *et al.* 2001); la presenza di ulteriori contaminazioni, di vettori BAC o *E. coli*, nelle sequenze rimanenti è stata testata mediante megablast.

Le sequenze degli inserti BAC sono state quindi assemblate mediante l'utilizzo del programma PHRAP (www.phrap.org).

2.2.3 Identificazione *in silico* dei geni NITR di *D. labrax*

Allo scopo di identificare i geni NITR all'interno del contiguo genomico di branzino, tutte le sequenze proteiche dei NITR depositate in *database* pubblici sono state raccolte e impiegate come *query* per analisi di Blast (Wolfsberg *et al.* 2001), utilizzando l'opzione *tblastn*.

Il *software* APOLLO (Lewis *et al.* 2002) è stato impiegato per visualizzare l'organizzazione della regione genomica analizzata e per annotare manualmente su di essa le strutture dei diversi geni NITR sulla base delle informazioni ottenute dai risultati di Blast. Il peptide segnale è stato predetto mediante l'utilizzo del *server* SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>) mentre il programma TMHMM v2.0

(<http://www.cbs.dtu.dk/services/TMHMM/>) è stato impiegato per identificare i domini transmembrana.

Le diverse famiglie di recettori NITR sono state infine definite, sulla base del grado di identità aminoacidica ($\geq 70\%$) dei domini V, mediante il programma MatGat (*MATrix Global Alignment Tool*; Campanella *et al.* 2003).

2.2.4 Amplificazione del cDNA e sequenziamento di trascritti *full-length*

Allo scopo di validare sperimentalmente le giunzioni esone-introne e le regioni transmembrana predette, nonché per individuare i motivi di più piccole dimensioni, per tutti i geni NITR appartenenti al *cluster* sono state amplificate e sequenziate le intere regioni codificanti. Un microgrammo (μg) dell'RNA estratto da rene cefalico è stato retrotrascritto a cDNA utilizzando l'enzima Superscript II (Invitrogen™). Coppie di *primer*, specifiche per ciascun recettore NITR di branzino, sono state disegnate sulla base delle sequenze geniche predette *in silico*; quando possibile, il *primer reverse* e il *primer forward* sono stati disegnati in modo da includere l'intera regione codificante del gene. Le coppie di *primer* specifiche per ciascun gene sono riportate in Appendice2A.

Un microlitro di cDNA, diluito 1:10, è stato utilizzato come template per la reazione di PCR (*Polymerase Chain Reaction*) adottando le seguenti condizioni di amplificazione:

Denaturazione iniziale	94°C per 2 minuti	
Denaturazione	94°C per 45 secondi	} 45 cicli
<i>Annealing</i>	60°C* per 30 secondi	
Estensione	72°C per 45 secondi	
Estensione finale	72°C per 5 minuti	

*Per ciascun gene, la temperatura (T°) di *annealing* è stata determinata in accordo alla T° di *melting* di ciascuna coppia di *primer*.

Per cinque geni NITR sono state effettuate anche RACE (*Rapid Amplification of cDNA Ends*), sia 5' che 3', per amplificare e sequenziare i trascritti *full-length*.

2.2.5 Analisi filogenetiche

Analisi filogenetiche sono state eseguite sull'intero *set* di recettori identificati in *D. labrax* e nelle altre specie di teleostei allo scopo di determinare il grado di divergenza dei NITR identificati in branzino come anche identificare potenziali orologi tra i geni appartenenti alle diverse specie.

Tutte le sequenze proteiche dei domini V sono state allineate utilizzando il programma ClustalW (Thompson *et al.* 1994), mentre CodonAlign v2.0 è stato utilizzato per eseguire l'allineamento nucleotidico delle sequenze codificanti sia i domini V che i domini I. Per queste analisi tutte le sequenze dei NITR identificati in zebrafish, pesce gatto, medaka, trota, fugu, tetraodon e *Sphaeroides nephelus* sono state incluse nell'allineamento.

Gli alberi filogenetici sono stati ricavati utilizzando il metodo della *Maximum likelihood* (ML) (Felsenstein, 2004); per le sequenze proteiche, il modello evolutivo più appropriato è stato desunto applicando il criterio di Akaike implementato nel *software* ProtTest v.1.4 (Abascal *et al.* 2005). Le analisi di ML sono state eseguite col programma Phyml 2.4.4 (Guindon *et al.* 2003) utilizzando il modello evolutivo JTT+I+G+F; la robustezza della topologia ottenuta è stata poi valutata eseguendo un'analisi di *bootstrap* con 500 repliche.

Per le sequenze nucleotidiche, l'analisi di ML è stata eseguita utilizzando il modello evolutivo GTR+I+G. Il programma TREEFINDER (Jobb *et al.* 2004) è stato poi utilizzato per valutare la probabilità di ipotesi filogenetiche alternative.

2.2.6 Analisi di pressione selettiva

Diversi *test* per valutare la presenza di siti soggetti a selezione positiva sono stati eseguiti comparando, per ciascuna posizione aminoacidica, il numero di sostituzioni non-sinonime (dN) al numero di sostituzioni sinonime (dS); un rapporto dN/dS (ω) maggiore di 1 è solitamente considerato indice di selezione positiva. Diversi approcci

sono stati utilizzati per calcolare questo rapporto. Per prima cosa sono stati valutati una serie di modelli implementati nel *software* PAML v4 (Yang 2007): modelli “neutrali”, che limitano ω a valori uguali o inferiori a 1 (modelli M1, M7 e M8a), sono stati comparati a modelli che ammettono anche la selezione positiva ($\omega > 1$; modelli M2 e M8) tramite il *likelihood ratio test* (LRT). I modelli adottati per le frequenze di ciascun codone, F3x4 e F61, hanno dato risultati pressoché identici; per questo motivo verranno riportati solo i risultati ottenuti utilizzando il modello F3x4, il più semplice dei due.

Una volta scelto di utilizzare il modello M8, gli aminoacidi soggetti a selezione positiva sono stati identificati utilizzando l’approccio *Bayes Empirical Bayes* (BEB; Yang *et al.* 2005); solo valori di probabilità maggiori di 0.95 sono stati considerati significativi.

Due ulteriori approcci, basati anch’essi sulla *Maximum Likelihood* e che contemplano alcune variazioni nel tasso di sostituzioni sinonime; il *Random Effects Likelihood* (REL) e il *Fixed Effects Likelihood* (FEL) (Kosakovsky Pond *et al.* 2005), sono stati invece implementati nel *server* DATAMONKEY (Pond *et al.* 2005).

Per concludere, un modello basato sui codoni (MEC), che prende in considerazione anche le proprietà fisico-chimiche degli aminoacidi al momento della stima del tasso di sostituzioni non-sinonime (Doron-Faigenboim *et al.* 2006), è stato implementato nel *server* Selecton (Stern *et al.* 2007).

2.2.7 Analisi di genomica comparativa

Tre proteine fiancheggianti il *cluster* di geni NITR in branzino; RBM4 (*RNA binding motif 4*), TRAF2-binding protein (*TNF Receptor-Associated Factor 2*), e FXR2 (*Fragile X mental retardation gene 2*), sono state utilizzate per verificare l’esistenza di relazioni sinteniche tra i NITR in *D. labrax* e le altre specie di teleostei. Allo scopo di identificare e localizzare questi geni in *D. rerio*, *G. aculeatus* e *O. latipes*, le più recenti versioni dei loro genomi, depositati in Ensembl (<http://www.ensembl.org/index.html>), sono state analizzate mediante TBLASTN.

2.2.8 Modelling comparativo del dominio V

I modelli teorici dei domini V di tre geni NITR di branzino (DI28, DI29 e DI30) sono stati generati *in silico* con il programma Modeller 9v4 (<http://www.salilab.org/modeller/>), applicando vincoli di simmetria per poter comporre dei dimeri.

L'approccio di *modelling* comparativo implementato in Modeller si basa sull'osservanza di condizioni spaziali ricavate dall'allineamento della sequenza di interesse con una o più strutture modello (Sali *et al.* 1993). Come prima cosa, le sequenze di branzino sono state impiegate in una ricerca, tramite FASTA nel *database Protein Data Bank* (PDB) (<http://www.rcsb.org/>), allo scopo di identificare domini NITR analoghi; successivamente, il server ClustalW (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>) è stato utilizzato per eseguire gli allineamenti dei domini V per ogni possibile combinazione “sequenza *target*/stampo candidato”. Le strutture dei geni NITR10 e NITR11 di *Ictalurus punctatus*, recentemente risolte (Cannon *et al.* 2008), sono state scelte come “stampo” in quanto mostravano la maggiore similarità con gli omologhi recettori di branzino.

Le impostazioni di *default* del programma Modeller sono state quindi impiegate per generare cinque possibili modelli per ciascun gene NITR; il modello migliore è stato poi scelto sulla base delle sue proprietà sterico-chimiche.

2.2.9 Analisi di espressione dei geni NITR in *D. labrax*

I livelli di espressione di tutti i geni NITR identificati in questo studio sono stati esaminati, tramite RT-PCR, in quattro esemplari di branzino; per ciascun individuo sono stati analizzati 8 diversi tessuti *target*: branchie, milza, intestino, muscolo scheletrico, pelle, rene cefalico, fegato e sangue.

Per poter effettuare un'analisi semi-quantitativa dei livelli di espressione degli mRNA di ciascun recettore, il gene codificante la β -actina, noto *housekeeping*, è stato sempre incluso in ciascuna reazione di amplificazione, in qualità anche di controllo positivo.

Le dimensioni degli ampliconi ottenuti per ciascun gene sono stati valutati mediante elettroforesi su gel di agarosio (1,8%) contenente *SYBR Safe™ DNA Gel stain 10,000X* (Invitrogen™). I dati di espressione per ciascun campione sono stati formulati in sistema binario come presenza/assenza, test di correlazione non-parametrica (test dei ranghi di Spearman) sono stati inoltre eseguiti per ciascun confronto a coppie allo scopo di verificare se i profili di espressione dei diversi geni NITR rispecchiassero o meno le reciproche relazioni filogenetiche/evolutive.

2.3 RISULTATI

2.3.1 Identificazione di un cluster genico di NITR in *D. labrax*

Il sequenziamento *shot-gun* dei due BAC ipotizzati contenere geni NITR ha prodotto 2.021 sequenze (*reads*), di lunghezza media di circa 650 bp, per il clone bassbac-18k1, e 1.979 *reads*, di lunghezza media di poco superiore, 675 bp, per bassbac-79e10.

L'assemblaggio delle sequenze di bassbac-18k1 ha prodotto due contigui (rispettivamente di 136.804 bp e 10.268 bp), con una copertura media di oltre 8 volte, ordinati poi in un unico *scaffold* in base alle sequenze delle BAC-ends. Per bassbac-79e10, invece, l'assemblaggio ha generato 4 contigui (rispettivamente di 3.184 bp, 29.024 bp, 100.505 bp e 4.881 bp), con una copertura media di quasi 10 volte; anche questi sono stati ordinati in un unico *scaffold* unendo le informazioni delle sequenze BAC-*end* a quelle derivanti dalle *reads* dei plasmidi.

I due BAC ottenuti mostravano una regione sovrapposta di circa 13.5 kb e sono stati quindi allineati per comporre un unico *scaffold* di 271.448 bp (*gaps* incluse); il contenuto medio in GC di questa regione è del 40,5%.

L'analisi *in silico* di questo contiguo genomico ha permesso l'identificazione di un *cluster* genico di *Novel Immune-type receptor* in *D. labrax*.

Ricerche di Blast, utilizzando tutte le sequenze proteiche pubbliche dei geni NITR in altre specie di teleostei, hanno identificato 27 potenziali NITR (DI1-DI27), localizzati in entrambe le orientazioni trascrizionali. Tre ulteriori geni (DI28-DI30) sono stati individuati in un secondo momento sulla base della loro similarità ai primi NITR identificati in branzino (Figura 2.1).

Utilizzando l'approccio bioinformatico, è stata possibile l'identificazione solo di porzioni dei domini V e I, della regione transmembrana e di parte della coda citoplasmatica.

Allo scopo di confermare sperimentalmente le predizioni *in silico*, nonché individuare le porzioni geniche rimanenti, il *full-length* di 5 trascritti (DI8, DI10, DI20, DI24 e DI25), nonché le intere regioni codificanti di tutti gli altri membri del cluster, sono stati amplificati e sequenziati.

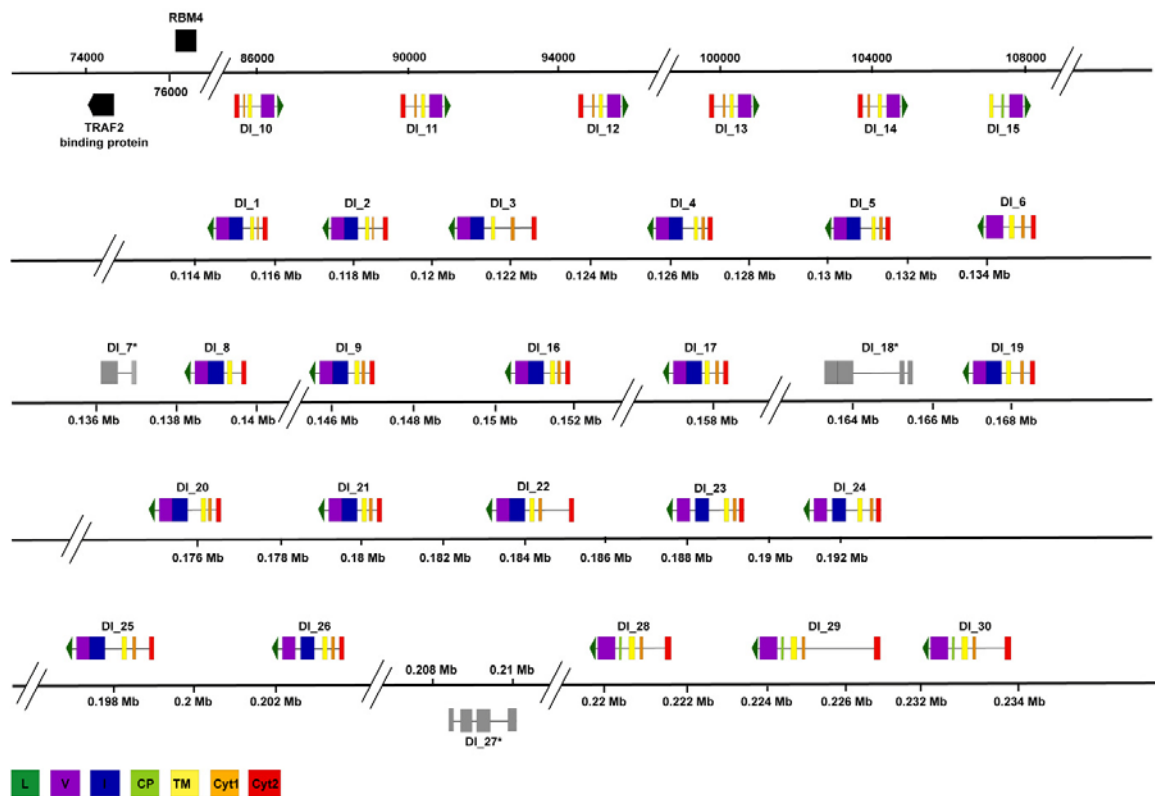


Figura 2.1: Organizzazione genomica del *cluster* di geni NITR identificati in *D. labrax*. In corrispondenza di ciascun gene ne viene indicato il nome, la direzione delle punte indica, invece, l'orientazione trascrizionale. I rettangoli corrispondono agli esoni. (L= Sequenza *leader*, V: Dominio V, I: Dominio I, TM: Regione transmembrana, Cyt1 and Cyt2: Domini citoplasmatici, CP: potenziale peptide di connessione). In grigio sono indicati i potenziali pseudogeni.

Tre dei NITR inizialmente identificati sono risultati essere costituiti da porzioni di domini sparsi e non hanno mai dato prodotto in seguito ad amplificazione del cDNA; per questo motivo questi geni (DI7, DI18 e DI27) sono stati considerati pseudogeni.

Sia la sequenza del contiguo genomico che quelle di tutti i geni NITR del *cluster* sono stati depositati in GenBank con *accession number* rispettivamente GQ385066 e da GQ385039 a GQ385065.

2.3.2 I geni NITR condividono elementi conservati

Come già riportato per altre specie di teleostei, i NITR di branzino risultano essere correlati alle Ig e ai TCR mediante caratteristiche conservate. I residui canonici del dominio V; G¹⁶, C²³, W⁴¹, G⁴⁷, L⁸⁹, D⁹⁸, A¹⁰⁰, Y¹⁰², C¹⁰⁴ e A¹⁰⁵, che sono comuni tra Ig e TCR, sono stati infatti riscontrati sia nei domini V che, in parte, nei domini I dei NITR (Figura 2.2). Otto residui di cisteina (2 nel dominio V e 6 nel dominio I), probabilmente coinvolti nella creazione dei ponti disolfuro come avviene nelle Ig, sono stati identificati in tutti i geni NITR. Motivi *Joining* (J; FGXGTXLXV) o *J-like*, localizzati nella porzione C-terminale del dominio V o del dominio I, sono anch'essi una caratteristica conservata in quasi tutti NITR di branzino (Vedi Appendice 2B).

Nonostante alcuni elementi risultino essere comuni a tutti i membri di questo *cluster* genico, un totale di soli 8 residui nel dominio V e 22 residui nel dominio I risultano essere assolutamente conservati in tutti i geni NITR di *D. labrax* (Figura 2.2).

Tutti i NITR identificati in branzino finora possono essere classificati come inibitori in base alla presenza di uno o due motivi ITIM (o *ITIM-like*) nella coda citoplasmatica, nonché la mancanza di residui carichi positivamente nella regione transmembrana. L'unica eccezione è rappresentata da DI15, nel quale il dominio citoplasmatico è completamente assente. Nessun membro del *cluster*, inoltre, presenta motivi attivatori come le sequenze ITAM-correlate o ITSM (Yoder 2009). È interessante notare, invece, come la distanza tra i residui di tirosina dei domini ITIM e quelli delle regioni *ITIM-like* sia sempre di 23-25 aminoacidi (vedi Appendice 2B), del tutto simile a quanto riscontrato anche nei KIR (Bléry *et al.* 2000).

In accordo con la conservazione complessiva dei principali motivi strutturali, una forte correlazione tra i NITR di branzino e gli omologhi identificati in altre specie di teleostei è stata ulteriormente confermata dall'analisi della struttura secondaria di questi recettori.

Mediante un approccio di *modelling* comparativo, la similarità maggiore è stata di fatto riscontrata con le strutture dei NITR10 e NITR11 di *I. punctatus* (Cannon *et al.* 2008), già dimostrate essere strutturalmente simili ai TCR sia nella topologia che nel sistema di dimerizzazione.

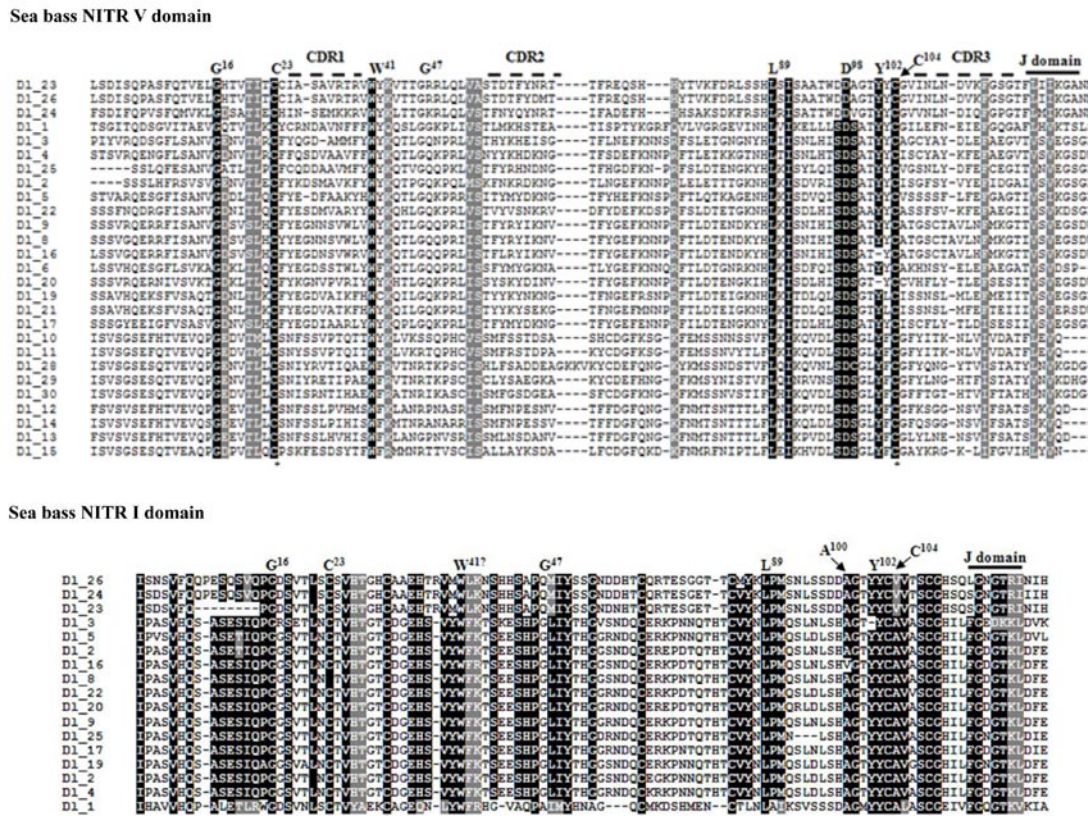


Figura 2.2: Allineamento delle sequenze proteiche dei domini V e I di *D. labrax*. I residui identici sono evidenziati in nero, i residui simili sono evidenziati in grigio. Sono indicate le posizioni delle regioni corrispondenti alle CDR1, CDR2 e CDR3 come anche dei domini J-simili. I residui canonici, comuni tra Ig e TCR, vengono indicati dalle lettere corrispondenti, i numeri si riferiscono alle posizioni designate dall'*International Immunogenetics Database* (IMGT) (Lefranc *et al.*, 2003). Le cisteine conservate sono indicate da un asterisco (*) sotto l'allineamento.

2.3.3 Il cluster genico dei NITR è costituito da diverse famiglie di recettori

Il confronto delle sequenze proteiche predette dei NITR di *D. labrax* mostra un elevato grado di variabilità. I 27 geni NITR identificati possono essere suddivisi in 18 diverse famiglie sulla base del grado di identità del dominio V, è convenzionalmente riconosciuto, infatti, che una famiglia genica comprenda tutte le sequenze con identità aminoacidica >70%.

Come già riportato per i geni NITR di altre specie di teleostei, queste famiglie differiscono nel numero di domini extracellulari, nella loro organizzazione genomica e per la presenza/assenza, nonché la localizzazione, dei motivi J, J-like, ITIM e ITIM-like. Un'analisi evolutiva dei geni NITR di branzino è stata condotta utilizzando le sequenze proteiche del dominio V, l'unico ad essere presente in tutti i membri del *cluster* genico. L'analisi filogenetica ha suddiviso i NITR in tre gruppi principali, supportati da un elevato valore di *bootstrap* (Figura 2.3).

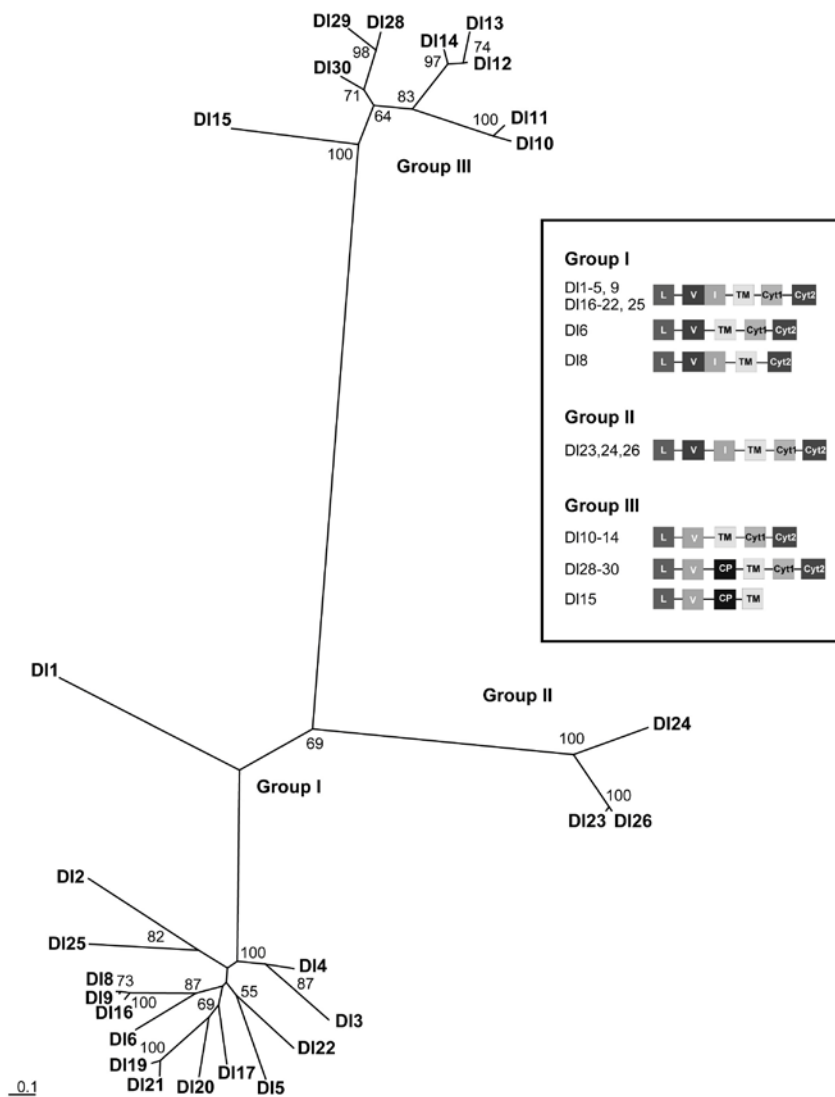


Figura 2.3: Albero filogenetico di domini V codificati dai 27 NITR di *D. labrax*. A ciascun nodo è stato assegnato il rispettivo valore di *bootstrap*; i valori al di sotto di 50 non sono stati riportati. La lunghezza di ciascun ramo corrisponde al numero di sostituzioni aminoacidiche (la scala è indicata in basso a sinistra). A destra è indicata anche l'organizzazione genomica di ciascun gruppo di NITR (non in scala). I rettangoli corrispondono agli esoni (L= Sequenza leader, V: Dominio V, I: Dominio I, CP: Potenziale peptide di connessione, TM:Regione transmembrana, Cyt1 e Cyt2: Domini citoplasmatici)

Gruppo I: è costituito di 15 geni (D11-6, D18, D19, D116, D117, D119-21, and D125) nei quali il dominio V e il dominio I sono codificati da un unico esone. Tutti i NITR appartenenti a questo gruppo sono inoltre caratterizzati dalla presenza di 2 motivi ITIM nella regione citoplasmatica e dei motivi GXG e FGXG nelle regioni C-terminali rispettivamente dei domini V e I (vedi Appendice 2B). Gli unici membri del Gruppo I a non conformarsi a queste caratteristiche comuni sono D16, che manca del dominio I, e D18, nel quale il dominio Cyt1 è assente.

Gruppo II: in questo gruppo, costituito di soli 3 geni (D123, D124 e D126), il dominio V e il dominio I sono codificati da due esoni distinti. Anche in questo caso i NITR presentano uno o due motivi ITIM nella regione citoplasmatica e le sequenze J sono anche qui presenti ma posizionate in ordine inverso rispetto al Gruppo I ([GXG] nel dominio V e [FGXG] nel dominio I).

Gruppo III: i geni appartenenti a questo gruppo (D110-15 e D128-30) sono caratterizzati dalla presenza di un solo dominio extracellulare, il dominio V, mentre il dominio I è assente in tutti. Sono assenti anche le sequenze J o *J-like* mentre viene ulteriormente confermata la presenza dei domini ITIM nella coda citoplasmatica (l'unica eccezione è rappresentata da D115, troncato a livello della regione transmembrana). All'interno di questo gruppo, 4 geni (D115, D128-30) sono caratterizzati dalla presenza di un ulteriore esone, di dimensioni comprese tra le 30 bp e le 42 bp, posizionato tra il dominio V e la regione transmembrana. Questo esone non risulta essere allineabile tra D115 e D128-30, né a livello nucleotidico né a livello aminoacidico, suggerendo che possa essere stato acquisito in questi due gruppi in modo indipendente. Il *modelling* del gene D128 suggerisce che questo esone possa codificare un peptide di connessione ("*stalk*") tra l'unico dominio extracellulare e la regione transmembrana. Fino ad ora, una simile organizzazione è stata osservata solo nei NITR5-11 del pesce gatto.

2.3.4 I NITR seguono il modello evolutivo “*birth-and-death*”

Allo scopo di analizzare i NITR di branzino da una prospettiva evolutiva più ampia, un albero filogenetico, basato sull'allineamento del dominio V, è stato costruito considerando tutte le sequenze di NITR pubbliche (Figura 2.4). Nodi supportati da valori di *bootstrap* >50% (sebbene nella maggior parte dei casi inferiori al 90%) hanno permesso di identificare diversi *cluster* genici; in Figura 2.4 questi nodi sono indicati con una lettera dell'alfabeto per renderne più agevole l'osservazione.

I tre gruppi più basali (nodi A, B e C), che contengono quasi esclusivamente NITR di zebrafish, pesce gatto e trota, non includono alcun gene di branzino. Tutti i geni identificati in *D. labrax* sono distribuiti in diversi *cluster* tutti però divergenti a partire dal nodo D. Un primo gruppo (nodo E, supporto di *bootstrap* 85%) contiene esclusivamente NITR formati da un unico dominio extracellulare (dominio V); all'interno di questo, un sottogruppo (nodo F, supporto di *bootstrap* 99%) è costituito unicamente da specie appartenenti all'ordine degli *Acanthopterygi* (medaka, tetraodon, branzino). Tuttavia, non tutti i NITR costituiti dal solo dominio V (da ora in poi definiti *V-only*) sono inclusi nel *cluster* E. I NITR5-11 del pesce gatto (Ip5-11), tutti mancanti del dominio I, sono concentrati in un gruppo a parte (divergente al nodo A).

Il nodo G raccoglie il più ampio insieme di geni, incluso un primo *cluster* (nodo H) che contiene esclusivamente geni nei quali i domini V e I sono codificati da 2 esoni separati, ed un secondo gruppo (nodo J) che contiene principalmente NITR con domini V ed I codificati da un unico esone. La maggior parte dei geni che mostrano quest'ultima organizzazione genomica sembrano avere un'origine comune (nodo N), sebbene siano osservabili 2 eccezioni: NITR24 e NITR10 di medaka (O124 e O110), come anche NITR1 di *D. labrax* (D11), NITR10 di *S. nephelus* (Sn10) e un NITR di *T. nigroviridis* (Tn72801) (*cluster* L). Ipotesi filogenetiche alternative, che forzavano D11, Sn10, and Tn72801 all'interno del gruppo N, sono state testate e statisticamente rigettate.

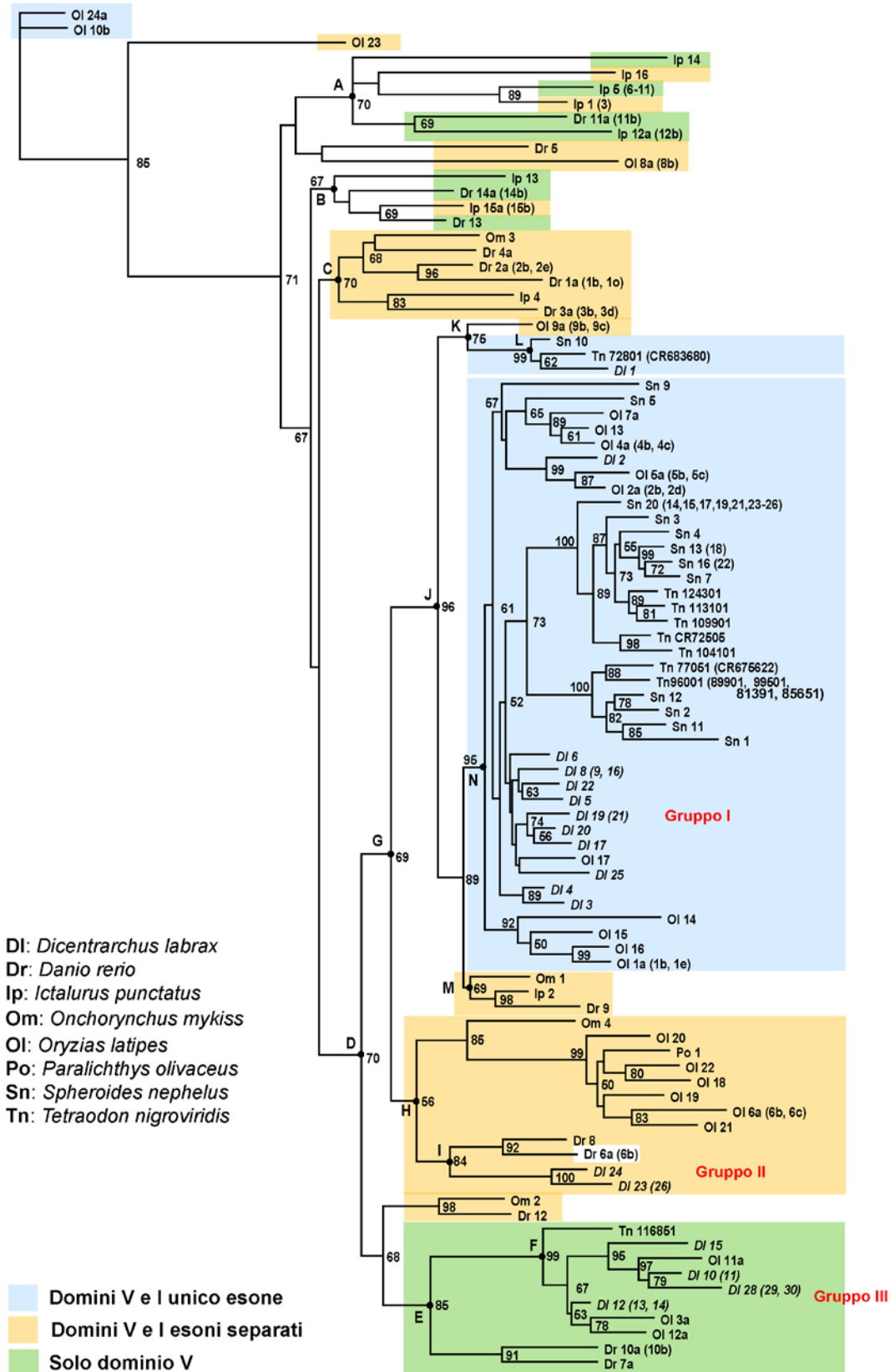


Figura 2.4: Albero filogenetico che mostra le relazioni evolutive tra i domini V di tutti i NITR pubblici. Ciascun gene è definito dall'abbreviazione della specie a cui appartiene seguita dal numero attribuitogli nella specie stessa. Nell'albero è stato incorporato un solo membro per ciascuna famiglia, gli altri geni appartenenti alla stessa famiglia sono indicati in parentesi. A ciascun nodo è stato assegnato il rispettivo valore di *bootstrap*; i valori al di sotto di 50 non sono stati riportati.

I geni NITR che mostrano questa organizzazione genomica sono stati individuati finora solo in specie appartenenti all'ordine degli *Acanthopterigi*, nelle quali questi rappresentano la maggior parte dei geni identificati (23 su 23 in *S. nephelus*, 22 su 44 in *O. latipes* e 13 su 27 in *D. labrax*).

Un'analisi filogenetica più dettagliata, basata sull'allineamento nucleotidico delle regioni codificanti sia il dominio V che il dominio I, mostra che con la sola eccezione dei precedentemente citati OI24, OI10, D11, Sn10 e Tn72801, tutti questi geni sembrano essersi evoluti tramite indipendenti espansioni specie-specifiche. (Appendice 2C).

2.3.5 I geni NITR sono soggetti a selezione positiva

Un serie di test statistici diversi sono stati impiegati allo scopo di verificare se diverse espansioni geniche siano soggette a pressione selettiva; quattro gruppi di geni NITR sono stati presi in considerazione: il gruppo di NITR “V-only” di branzino (Gruppo III in Figura 2.3), e le tre espansioni geniche specie-specifiche (*D. labrax*, *O. latipes*, *S. nephelus*) caratterizzate dai domini V ed I codificati da un unico esone (Espansioni “Intronless”, riportate in Appendice 2C).

I *Likelihood Ratio Tests* (LRT; vedi Paragrafo 2.2.5) hanno fornito chiare evidenze, corredate da una elevata significatività statistica, di selezione positiva per tutti i quattro gruppi di geni considerati, in quanto il modello di evoluzione “neutrale” (M1), nonché i modelli ristretti solo a selezione negativa o neutrale (M7 e M8a), si sono rivelati inadatti rispetto al modello che prevede anche l'intervento di selezione positiva (M8).

Quattro diversi approcci sono stati quindi utilizzati per identificare i codoni (e quindi gli aminoacidi) soggetti a selezione positiva. Per poter applicare più metodi, viene comunemente suggerito di utilizzare un valore di α piuttosto elevato (almeno 0,25) seguito da un approccio basato sul consenso dei risultati ottenuti (Kosakowsky *et al.* 2005). In questo studio, anche utilizzando un valore di α ancora più conservativo (0,05)

e considerando solo i siti predetti come soggetti a selezione positiva in almeno 2 o 3 metodi su 4 (50% o 75% di *Majority Consensus* (MC)), diverse posizioni aminoacidiche sono risultate essere soggette a cambiamenti adattativi. I risultati ottenuti da tutti e 4 i metodi utilizzati sono riportati, per ciascun gruppo di geni, in Appendice 2D.

Poiché non tutti i recettori appartenenti al gruppo *V-only* di branzino possiedono la coda citoplasmatica, l'analisi si è limitata solo al Peptide segnale e al dominio V; tuttavia, è stato possibile identificare otto posizioni soggette a selezione positiva (75% di MC), di cui una confermata da tutti i quattro approcci utilizzati. E' importante sottolineare che sette di questi siti sono localizzati nelle *Complementarity Determining Region* (CDR1, CDR2, CDR3 in Figura 2.5); l'ottavo residuo, supportato da tutte e 4 le analisi, è immediatamente adiacente alla CDR2. In NITR28, una inserzione di 4 aminoacidi (indicata dalla freccia in Figura 2.5) separa questo sito dalla CDR2, provocando una modificazione sostanziale della struttura proteica predetta e determinando quindi un'ulteriore diversificazione dei recettori appartenenti al gruppo *V-only*.

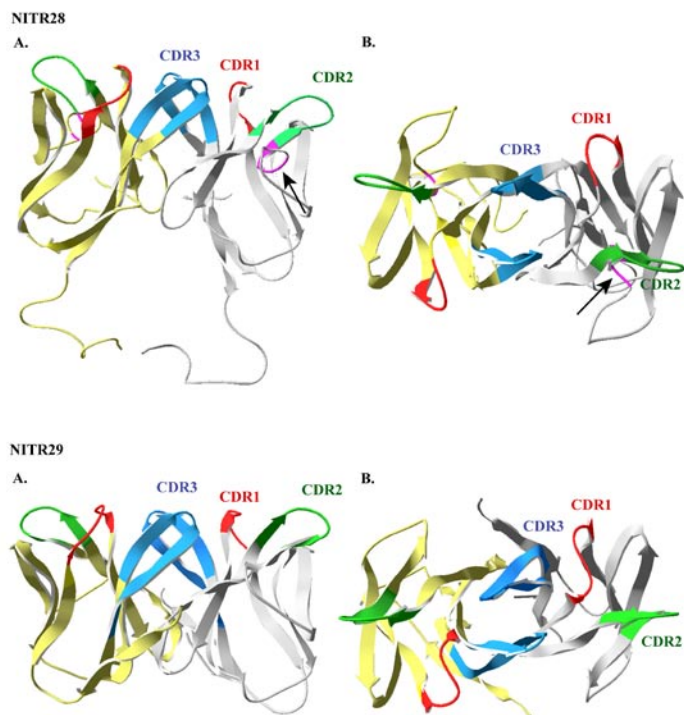


Figura 2.5: Modello tridimensionale degli omodimeri costituiti dai domini V di NITR28 e NITR29 di *D. labrax*. La struttura secondaria è raffigurata in giallo per una subunità, in grigio per l'altra. La CDR1 è evidenziata in rosso, la CDR2 in verde e la CDR3 in blu. Le frecce indicano l'inserzione di quattro aminoacidi in corrispondenza della CDR2 del NITR28. Viene raffigurato anche il peptide di connessione di NITR28.

In proporzione, un numero inferiore di posizioni aminoacidiche sono state indicate come soggette a selezione positiva nei rimanenti tre gruppi. Considerando di regola il 75% di MC, 4 siti sono stati identificati nell'espansione "Intronless" di *D. labrax* (29 con il 50% di MC), 3 in quella di *S. nephelus* (22 con il 50% di MC) e solo 2 nell'espansione genica di *O. latipes* (6 con il 50% di MC). Anche in questi gruppi di geni i siti maggiormente supportati sono risultati essere quelli localizzati nelle CDR (Appendice 2D), altri domini (Cyt2 e dominio I) sono comunque ben rappresentati.

Tabella 2.1: Siti aminoacidici dei NITR di branzino appartenenti al Gruppo III ("V-only") identificati essere sotto selezione positive.

AA position ^a	Protein domain ^b	Maximum Likelihood Method			
		PAML ^c	FEL ^d	REL ^e	MEC ^f
7	Signal P.		0.15		
9	Signal P.		0.12		
46	V_CDR1	0.993	0.02	804.9	
48	V_CDR1		0.04	97.3	
51	V	0.984		59.2	
55	V		0.04	42.5	
56	V		0.10	35.8	
69	V_CDR2			27.9	
70	V_CDR2		0.10		
71	V_CDR2	0.987	0.16	112	
72	V_CDR2	0.988			
73	V_CDR2	0.997		353.6	5.4
79	V	1.00	0.01	3763.6	5.2
80	V		0.12		
93	V		0.19		
95	V	0.984	0.01	594	
101	V	0.987		32.7	
117	V_CDR3		0.17		
118	V_CDR3	0.999	0.09	30	
119	V_CDR3	0.996		184.2	6
120	V_CDR3	0.978		62.1	4.2
121	V_CDR3			31.7	
127	V_CDR3		0.09	21.4	
130	V		0.10	23.7	

^a La posizione aminoacidica si riferisce all'allineamento dei NITR appartenenti al Gruppo III riportato in Appendice 2B.

^b V: dominio V, V_CDR1: *Complementarity determining region 1*, V_CDR2: *complementarity determining region 2*, V_CDR3: *Complementarity determining region 3*, V/C2: dominio I, TM: regione transmembrana, Cyt1 e Cyt2: domini citoplasmatici.

^c Probabilità associate all'analisi con BEB.

^d *P-value*.

^e *Bayes factor*.

^f Intervallo di confidenza.

Tabella 2.2: Siti aminoacidici dei NITR di branzino appartenenti al Gruppo I (espansione “*Intronless*”) identificati essere sotto selezione positive.

AA position ^a	Protein domain	Maximum Likelihood Method			
		PAML	FEL	REL	MEC
4	Signal P.		0.11	25.7	
6	Signal P.		0.13	24.6	
10	Signal P.		0.20		
17	Signal P.		0.11	36.6	
30	V		0.04	52.6	
66	V_CDR1		0.09	39.2	
72	V		0.07	59.4	
83	V_CDR2		0.06	20.9	[1, 3.1]
88	V_CDR2		0.17		
94	V		0.07	67.5	
120	V		0.16		
127	V		0.15		
133	V_CDR3		0.05	33.8	
136	V_CDR3	0.998			[3.1, 3.1]
140	V_CDR3	0.994			[1, 3.1]
143	V		0.06	81	
144	V		0.04	103.7	
148	V		0.17	20.6	
151	V		0.10	41.6	
209	V/C2		0.20		
218	V/C2		0.12	31.1	
234	V/C2		0.17		
246	V/C2		0.19		
264	TM		0.12	22.8	
266	TM		0.05	60.5	
272	TM		0.11	24.6	
275	TM		0.13	22.6	
286	TM		0.12	23.1	
291	TM	0.978	0.08		
292	TM		0.03	63	
295	TM	0.972			
300	TM		0.16	22.7	
302	TM		0.16	20.3	
323	Cyt1		0.20		
338	Cyt2	0.981	0.12		[1, 3.1]
340	Cyt2		0.13	28.6	
348	Cyt2	0.997	0.06		[3.1, 3.1]
350	Cyt2	1.00	0.02		[3.1, 3.1]
351	Cyt2		0.07	30.6	

^a La posizione aminoacidica si riferisce all’allineamento dei geni appartenenti al Gruppo I riportato in Appendice 2B.

^b V: dominio V, V_CDR1: *Complementarity determining region 1*, V_CDR2: *complementarity determining region 2*, V_CDR3: *Complementarity determining region 3*, V/C2: dominio I, TM: regione transmembrana, Cyt1 e Cyt2: domini citoplasmatici.

^c Probabilità associate all’analisi con BEB.

^d *P-value*.

^e *Bayes factor*.

^f Intervallo di confidenza.

2.3.6 I NITR di branzino mostrano relazioni sinteniche con altre specie di teleostei

Un'analisi genomica comparativa è stata effettuata mediante analisi di Blast contro i genomi di *O. latipes*, *D. rerio*, e *G. aculeatus* allo scopo di identificarne le regioni che mostrassero dei *match* significativi con i geni NITR di branzino.

Le regioni identificate sia in zebrafish che in medaka corrispondono a quelle già descritte rispettivamente da Yoder e collaboratori (2002; 2004) e da Desai e colleghi (2008); le analisi condotte in questo studio hanno inoltre permesso di identificare i potenziali ortologi di FXR2, RBM4, e TRAF2-binding protein (Vedi paragrafo 2.2.6): i tre geni risultano essere collocati in prossimità del *cluster* di geni NITR nel chr. 18 di medaka e in quello localizzato nel chr. 7 di zebrafish (Figura 2.6).

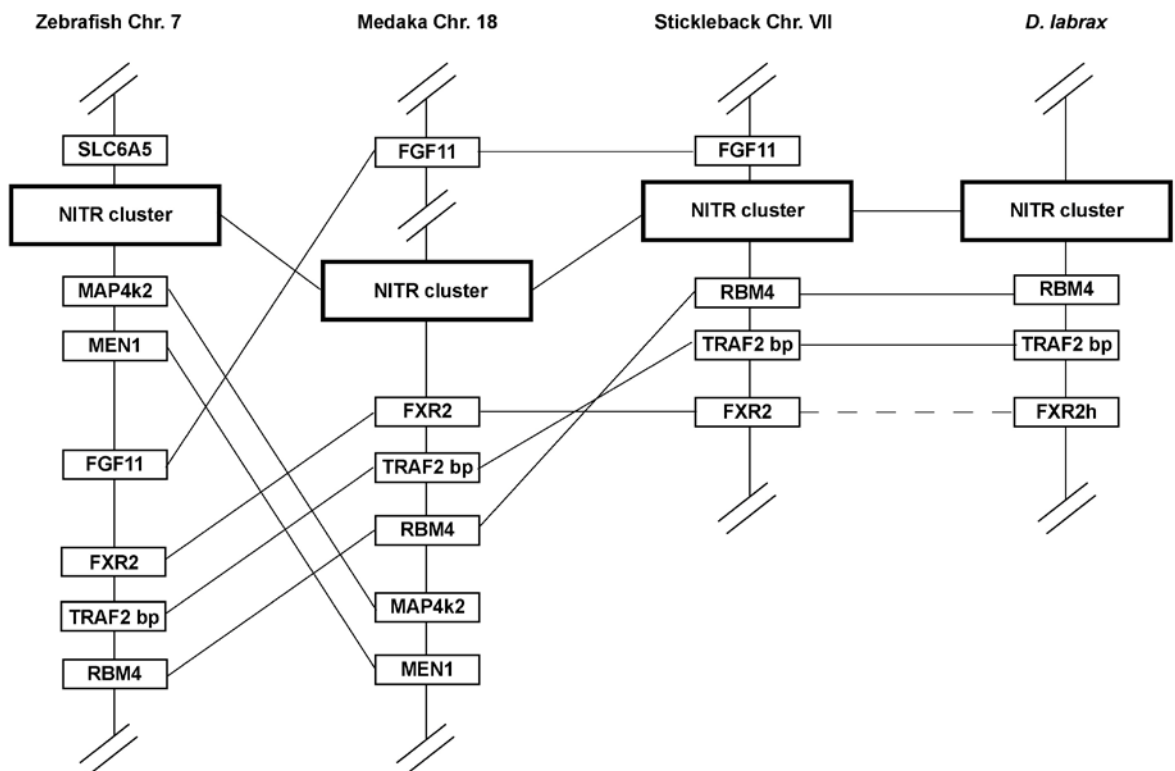


Figura 2.6: Confronto tra il cluster di geni NITR in *D. labrax* con le regioni sinteniche di zebrafish, medaka e stickleback. I geni (rettangoli) e le distanze tra di essi non sono in scala.

Oltre a caratterizzare i recettori NITR in medaka, Desai e collaboratori (2008) hanno individuato un *cluster* genico anche nel gruppo (Grp.) IV di stickleback; in questo studio, un ulteriore complesso di geni NITR, che copre un totale di tre contigui genomici (contig_7245-7247, ENSEMBL release 54), è stato identificato nel Grp. VII di *G. aculeatus*. Questo *cluster* si estende per circa 250 kb ed è costituito di almeno 24 NITR caratterizzati dalla presenza di domini V ed I codificati da uno o due esoni. I geni codificanti FXR2, RBM4, e l'omologo di TRAF2-binding protein sono stati inoltre identificati al 3' di questa regione genomica, indicando che il *cluster* di geni NITR di branzino e quello di stickleback localizzato nel Grp. VII hanno una relazione sintenica con quelli posizionati nel chr. 18 di medaka e il chr. 7 di zebrafish (Figura 2.6).

2.3.7 Analisi dell'espressione genica dei NITR di *D. labrax*

Una prima osservazione dei livelli di espressione genica dei NITR di branzino è stata effettuata, mediante RT-PCR *end-point*, su tessuti prelevati da quattro diversi esemplari giovanili. Sia organi immuno-competenti che non immuno-competenti sono stati utilizzati per la retrotrascrizione del mRNA e *primer* gene-specifici sono stati poi impiegati nella reazione di amplificazione. I dati di espressione di ciascun gene per individuo e tessuto, formulati come presenza/assenza, sono riportati in Appendice 2E.

Per tre geni; D12, D14 e D116, sono stati rilevati livelli di espressione nulli o molto deboli mentre per tutti gli altri membri del *cluster* sono state osservate variazioni di espressione sia tra individui che tra i diversi tessuti analizzati.

Come prevedibile, buoni livelli di espressione dei geni NITR vengono riscontrati principalmente nei tessuti che giocano un ruolo rilevante nella risposta immunitaria; tutti i geni NITR risultano, infatti, espressi nel rene cefalico e, con poche eccezioni, nelle branchie, nell'intestino e nella milza.

Una debole espressione, e variabile tra i diversi individui, è stata invece osservata nel sangue (D113-15, D119, D121, D114 and D126), nel fegato (D11, D113, D115, D119, D121-23, D125 and D126) e nel muscolo scheletrico (D11, D13, D15, D18, D111-15, D121-23, D125 and D128).

Nell'ambito dell'espansione/duplicazione genica, un dibattito tuttora aperto riguarda la possibilità che i profili di espressione genica vengano mantenuti anche nelle copie dei geni duplicati. Test di correlazione non parametrica eseguiti sui dati di espressione raccolti per i NITR di branzino (in termini di presenza/assenza) mostrano che, per alcune coppie di geni, i livelli di espressione risultano essere positivamente correlati, suggerendo che in molti casi la regolazione trascrizionale verrebbe mantenuta anche dopo la duplicazione. La correlazione tra i profili di espressione, tuttavia, non sempre riflette le relazioni filogenetiche che intercorrono tra i diversi geni NITR; un esempio di questo può essere considerato il gruppo di geni D110-14. Sulla base dell'albero filogenetico riportato in Figura 2.3 e della loro organizzazione genomica (Figura 2.1), questi recettori sembrerebbero avere origine da un unico gene. Una prima duplicazione avrebbe dato origine agli antenati di D110/11 e di D112/13/14, seguita da una seconda duplicazione, che ha determinato la generazione di D110 e D111 come anche D114 e l'antenato di D112/13; un ultimo evento poi sarebbe deputato della creazione di D112 e D113. In accordo con la loro distanza filogenetica, il confronto dell'espressione di D110 e D111 mostra un'identità dell'84% (27 su 32) tra le diverse combinazioni esemplare/tessuto, con un coefficiente di correlazione (rho di Spearman) di 0,74 ($p < 0.00001$). Al contrario, nonostante anche le sequenze di D112 e D113 siano strettamente correlate, D113 sembra essere soggetto a una diversa pressione regolatoria; questi due geni, infatti, mostrano una identità di espressione solo del 56% (rho di Spearman = 0.24, $p = 0.17$).

2.4 DISCUSSIONE

I geni NITR sono stati ipotizzati essere gli omologhi funzionali dei *Natural Killer Receptor* (NKR) nei mammiferi, i quali possono inibire o attivare la citotossicità e/o il rilascio di citochine da parte delle cellule NK (Yoder 2009). In comune con gli altri geni del sistema immunitario, sia gli NKR che i NITR sembrano seguire il modello evolutivo noto con il nome di “*birth-and-death*” (Nei *et al.* 2005). Questo modello assume che quando si confronta una famiglia multi genica tra due specie, diversi geni appaiono avere avuto origine da un *set* di paraloghi, presenti nell’antenato comune, mediante una serie di duplicazioni in *tandem* (fenomeno noto come “*birth*”) verificatesi in seguito alla separazione delle due linee filitiche. D’altra parte, alcune copie possono essere state invece perse in seguito a delezioni o fenomeni di pseudogenizzazione (fenomeno noto come “*death*”).

Il modello evolutivo “*birth-and-death*” prevede quindi la presenza in ciascuna specie di un mix tra geni ortologhi (presenti nell’antenato comune e ritenuti in singola copia in entrambe le specie), alcuni co-ortologhi (geni che sono andati incontro ad espansione in una o entrambe le specie mediante duplicazioni specie-specifiche) e geni paraloghi (geni duplicati presenti nell’antenato comune e che sono andati persi in una delle due specie mentre sono stati ritenuti nell’altra). In contrasto con quanto si verifica in caso di evoluzione concertata, a seconda delle forze evolutive in atto su una specifica famiglia genica, tra i membri di quest’ultima è atteso un certo grado di divergenza a livello di sequenza, come anche la presenza di alcuni pseudogeni (Nei *et al.* 2005).

Conformemente a quanto previsto dal modello “*birth-and-death*”, anche nel *cluster* di geni NITR identificato nel branzino è stata riscontrata la presenza di tre pseudogeni (10% dei geni isolati), e, analogamente a quando osservato anche in altre specie di teleostei, sono state riconosciute 18 diverse famiglie e tre gruppi principali di recettori,

ciascuno caratterizzato da una distintiva organizzazione genomica e composizione in domini (Figura 2.3).

L'evoluzione e la diversificazione dei geni NITR in branzino appare essersi verificata mediante duplicazioni in *tandem* ripetute, in quanto molti dei geni risultano essere organizzati in piccoli sottogruppi genomici che anche fisicamente riflettono la loro posizione filogenetica (Figure 2.1 e 2.3). La corrispondenza che, infatti, si osserva tra la distanza filogenetica e quella fisica dei membri delle diverse famiglie geniche suggerisce che il meccanismo principale alla base della loro amplificazione possa essere un *crossing-over* ineguale. La famiglia di DI8, per esempio, include tre geni (DI8, DI9 e DI16) che potenzialmente derivano da un evento di duplicazione piuttosto recente; la percentuale di identità delle proteine codificate da essi, infatti, è superiore al 85%; DI9 e DI16, in particolare, codificano per proteine identiche al 95%, e anche le sequenze non codificanti (introni) mostrano un grado di identità nucleotidica maggiore del 90%. Una situazione simile è riscontrata anche nel caso della famiglia di DI28 (DI28-DI30) dove le sequenze codificanti sono identiche per l'85-91% e anche gli introni, a eccezione di uno, sono ancora allineabili.

Anche quando osservati da una più ampia prospettiva evolutivista, i diversi gruppi (Gruppo I-III) di NITR di branzino appaiono conformarsi al modello evolutivo "*birth-and-death*".

L'origine del Gruppo I può essere fatta risalire ad un *set* minimo di due soli geni, presenti nell'antenato comune degli *Acanthopterygi* (*O. latipes*, *S. nephelus*, *T. nigroviridis*, *D. labrax*); uno di questi due geni è stato poi ritenuto in singola copia in *S. nephelus*, *T. nigroviridis* e *D. labrax* (nodo L in Figura 2.4) mentre è andato perso in *O. latipes*; il secondo gene è invece andato incontro ad una serie di espansioni specie-specifiche (Nodo N in Figura 2.4 e, nel dettaglio, Appendice 2C).

Il Gruppo II è il meno numeroso in branzino e sembra che anche i geni omologhi nelle altre specie di teleostei (Nodo H in Figura 2.4) siano andati incontro ad un numero piuttosto limitato di eventi duplicativi.

Il Gruppo III è caratterizzato dall'assenza dell'esone codificante il dominio I che, in quattro geni, (D115, D128-30) sembra essere stato rimpiazzato da un nuovo esone di soli 10-14 aminoacidi. Sebbene i NITR "V-only" siano stati riportati in diverse altre specie (*I. punctatus*, *D. rerio* e *O. latipes*), l'acquisizione di un esone aggiuntivo è stata osservata finora esclusivamente in *I. punctatus*. Come già anticipato, questa acquisizione appare essersi verificata tre volte indipendentemente.

Due diversi meccanismi possono essere responsabili dell'acquisizione di nuovi esoni: il primo, noto con il nome di "exonization", richiede la formazione di siti di *splicing* adiacenti alla sequenza genomica neo-reclutata (Sorek, 2007) mentre il secondo, definito "exon shuffling", promuove l'acquisizione di un esone mediante *crossing-over* tra due geni non omologhi, determinando quindi l'inserimento di un esone, da un gene "donatore", all'interno di un pre-esistente introne del gene "ricevente" (Patthy, 2008).

In caso si fosse verificato il fenomeno di "exonization", all'interno del secondo introne dei geni NITR più strettamente correlati a D115 e D128-30, ma che non hanno acquisito l'esone addizionale, dovrebbe essere comunque riscontrata una sequenza simile al nuovo esone; non è questo il caso dei NITR di branzino, i geni DL10-14, infatti, non soddisfano questo requisito. L'"exon shuffling" sembra quindi essere il meccanismo che più probabilmente ha determinato l'acquisizione del nuovo esone nei NITR di branzino; ipotesi suffragata anche dal fatto che questo fenomeno è già stato documentato nei geni KIR (Rajalingam *et al.* 2004).

La funzione del peptide codificato da questo esone non è ancora chiara in quanto non sono stati identificati in esso né domini funzionali né motivi specifici. Il *modelling* (Figura 2.5) ottenuto per omologia dei geni NITR di branzino con NITR10 e NITR11

del pesce gatto *I. punctatus*, fa coincidere questo peptide con la regione “stalk” descritta per NITR11 (Hawke *et al.* 2001; Cannon *et al.* 2008); ulteriori analisi effettuate mediante il server TMHMM v2.0 ne confermano, inoltre, la localizzazione extracellulare. Le evidenze raccolte sembrano suggerire che questa piccola regione possa codificare un peptide di connessione, comunemente riscontrato in altre molecole Ig-correlate come i recettori NILT nei teleostei (Stet *et al.* 2005; Østergaard *et al.* 2008) e i KIR e ILT nei mammiferi (Torkar *et al.* 1998; Martin *et al.* 2000). Nei geni NILT è stato suggerito che questo peptide possa delimitare una tasca accessibile e quindi potenzialmente coinvolta nell’interazione con il ligando; rimane tuttavia da verificare se nei NITR l’acquisizione ripetuta di questo esone abbia un significato funzionale.

La convergenza evolutiva verso organizzazioni genomiche simili (numero di esoni, struttura esone-introne) è una caratteristica comune ai geni NITR delle diverse specie di teleostei. Al momento non è chiaro quale fosse l’organizzazione ancestrale di questo complesso di geni; tuttavia, se si ipotizza che in origine la struttura genomica dei NITR fosse caratterizzata da sei esoni e due domini Ig, allora le perdite dei domini I, TM e citoplasmatici, nonché dell’introne che separa le regioni codificanti i domini V e I, appaiono essersi verificate più volte indipendentemente.

Un’ulteriore evidenza della similitudine tra NITR e KIR è data dalla loro organizzazione modulare e la loro evoluzione per mezzo di diverse combinazioni di domini; anche nei recettori KIR, infatti, possono essere presenti un numero variabile di domini Ig (2 o 3) come anche code citoplasmatiche più o meno lunghe. Anche nel caso dei KIR, inoltre, sia gli alleli attivatori che inibitori del *locus* 3DL1/S1 sono risultati essere soggetti a selezione positiva in alcune popolazioni umane (Norman *et al.* 2007); le posizioni aminoacidiche sotto selezione sono posizionate principalmente nei due domini Ig (D1 e D2) che legano le molecole MHC di classe I come anche nel terzo dominio Ig (D0), che si suppone aumenti la loro capacità di legame.

In questo studio è emerso che quattro *cluster* di geni NITR, appartenenti a tre diverse specie (*D. labrax*, *O. latipes*, *S. nephelus*), sono stati oggetto di selezione positiva. Questa è la prima evidenza che l'elevata variabilità di questi recettori, riportata in diverse specie di teleostei, potrebbe essere dovuta, almeno in parte, a una pressione selettiva di tipo adattativo che ha favorito l'evoluzione di un gruppo di molecole altamente diversificate. Le espansioni geniche specie-specifiche sono una caratteristica ricorrente nei geni NITR di diverse specie (Yoder 2009); le informazioni raccolte in questo studio confermano ulteriormente il ruolo chiave di questo fenomeno in quanto duplicazioni ripetute all'interno di una stessa famiglia genica forniscono alle forze selettive la "materia prima" su cui agire. Questa osservazione è in accordo con diversi studi condotti su interi genomi allo scopo di definire il ruolo che la selezione positiva gioca nell'evoluzione genica; questi studi hanno infatti evidenziato che, tra i geni soggetti a questo genere di pressione selettiva, quelli che giocano un ruolo nel sistema immunitario risultano significativamente rappresentati (Bustamante *et al.* 2005; Kosiol *et al.* 2008).

L'utilizzo di modelli probabilistici basati sull'analisi dei codoni permette di identificare quali sono i siti di una specifica proteina più probabilmente soggetti a selezione positiva (Anisimova *et al.* 2009); sebbene le analisi condotte in questo studio non siano conclusive e necessitino di ulteriori approfondimenti, l'utilizzo di diversi metodi suggerisce che i codoni sotto pressione adattativa siano localizzati principalmente in regioni di interazione con il ligando (CDR1-3).

Uno studio recentemente condotto su medaka (Cannon *et al.* 2008) dimostra chiaramente come singole sostituzioni aminoacidiche a carico della CDR1 dei NITR di *I. punctatus* siano sufficienti ad acquisire/perdere la capacità di riconoscimento di determinati *target*. Anche la stretta somiglianza tra le strutture proteiche dei NITR e quelle dei TCR suggerisce che le CDR1-3 possano rivestire un ruolo analogo di

riconoscimento del ligando; non è quindi inaspettata l'individuazione di siti soggetti a selezione positiva proprio in queste regioni. E' comunque da tenere in considerazione che i modelli basati sull'analisi di codoni spesso sottostimano la variabilità tra le sequenze, in quanto sia delezioni che inserzioni non possono essere incorporate nell'analisi; questo ha portato, per esempio, all'esclusione di un'inserzione di quattro aminoacidi nella CDR2 di D128 (a ridosso del sito indicato come soggetto a selezione da tutti i modelli utilizzati, vedi Tabella 2.1), che, invece, ha un chiaro effetto nella struttura tridimensionale del dominio V (Figura 2.5).

Oltre alle regioni deputate al riconoscimento del ligando, ci sono evidenze di altri siti, localizzati nella porzione citoplasmatica, potenzialmente soggetti a selezione positiva. La funzione regolatoria di questa regione sembra essere conservata rispetto agli NKR (Yoder *et al.* 2001; Yoder *et al.* 2007; Wei *et al.* 2007), i segnali sia inibitori che attivatori sembrano essere, infatti, trasdotti per mezzo del legame con proteine intracellulari ad essi associate (ad es. DAP12, SHPs); è quindi possibile che modificazioni aminoacidiche a carico della coda citoplasmatica vadano ad alterare la specificità/affinità delle interazioni recettore/proteina, modulando quindi i segnali di trasduzione tra i diversi NITR.

Il riscontro di selezione Darwiniana nel dominio V dei NITR si accorda bene con l'ipotesi che i NITR siano gli omologhi funzionali dei KIR nei mammiferi, i quali mediano il riconoscimento del *self* da parte delle cellule NK. Si ipotizza, infatti, che i KIR inibitori siano soggetti a una forte pressione selettiva allo scopo di poter far fronte all'enorme varietà dei loro ligandi, codificati da centinaia di alleli nei loci MHC-I.

L'analisi comparativa di tutti i recettori NITR finora identificati fornisce ulteriori evidenze a favore del modello evolutivo "*birth-and-death*". Due *cluster* principali sono stati identificati nei cromosomi 7 e 14 di zebrafish (Yoder *et al.* 2001; Yoder *et al.* 2008), nei chr. 18 e 10 di medaka (Desai *et al.* 2008) e nei Grp. iV e VII di stickleback

(Desai *et al.* 2008; questo studio), che sono onologhi (coppie di cromosomi derivanti da un unico cromosoma ancestrale in seguito alla duplicazione dell'intero genoma; WGD: *Whole Genome Duplication*) (Nakatani *et al.* 2007). Come già affermato da Yoder e collaboratori (2009), queste osservazioni suggeriscono la presenza di un *set* di NITR nell'antenato comune di tutte le specie di teleostei, prima del verificarsi del WGD. In seguito a duplicazione, il *cluster* originale è stato mantenuto in duplice copia, nonostante singoli NITR siano stati poi alternativamente persi/duplicati/espansi indipendentemente nelle diverse specie di teleostei.

Per esempio, un'analisi approfondita dei NITR appartenenti al Nodo H (Figura 2.4) indica che i membri del Gruppo II di branzino sono correlati più strettamente con i NITR 6a,b e 8 di zebrafish (Dr6a, Dr6b, Dr8) che con i NITR 6a,b,c e 18-22 di medaka (Ol6a, Ol6b, Ol6c, Ol18-22), nonostante quest'ultima sia filogeneticamente più vicina al branzino rispetto a quanto lo sia zebrafish. Questi NITR di branzino e zebrafish, tuttavia, sono localizzati in *cluster* ortologhi, mentre quelli di medaka si trovano nel corrispondente onologo; questo indica che i geni ortologhi dei membri del Gruppo II di *D. labrax* sono stati mantenuti in zebrafish mentre in medaka sono presenti solo i corrispondenti paraloghi.

Tutte queste evidenze, assieme alle relazioni sinteniche riscontrate tra il *cluster* genico di branzino e i cromosomi IV/10/14 di stickleback/medaka/zebrafish, (Figura 2.6), suggeriscono che anche in branzino possa esistere un secondo cluster di geni NITR non ancora identificato.

Analogamente alle osservazioni riportate in altre specie di teleostei (Piyaviriyakul *et al.* 2007; Yoder *et al.* 2002), i geni NITR sono espressi principalmente nei tessuti che giocano un ruolo di rilievo nella risposta immunitaria mentre variazioni gene-specifiche sono state osservate negli altri tessuti. La debole, se non nulla, espressione osservata nel sangue troverebbe però una spiegazione nel fatto che, a causa della limitata quantità di

materiale biologico di partenza, i saggi di espressione non sono stati effettuati sulla frazione leucocitaria bensì direttamente sul sangue intero.

I dati di espressione presentati in questo studio sono del tutto preliminari e certamente necessitano di ulteriori studi di tipo quantitativo (mediante *Real-time* RT-PCR); tuttavia, una moderata correlazione tra coppie e gruppi di geni, in termini di presenza/assenza, è stata comunque individuata, suggerendo che in molti casi la regolazione trascrizionale viene parzialmente mantenuta anche in seguito a duplicazione genica. I profili di espressione, tuttavia, non rispecchiano perfettamente le relazioni filogenetiche che intercorrono tra i diversi NITR suggerendo, quindi, che possano agire anche regolazioni di altro genere che lasciano quindi spazio alla diversificazione dei geni NITR.



**3. SVILUPPO E VALIDAZIONE DI
UNA PIATTAFORMA *MICROARRAY*
PER *SPARUS AURATA***

3.1 INTRODUZIONE

Nonostante l'enorme sviluppo che l'acquacoltura ha avuto negli ultimi decenni sono ancora molti i problemi e le limitazioni che caratterizzano questo tipo di allevamento; per superare questi ostacoli, diverse lacune che ancora permangono sulla biologia delle specie ittiche necessitano di essere colmate. Una miglior comprensione dei meccanismi che stanno alla base di caratteristiche produttive chiave offre l'opportunità di rivoluzionare l'acquacoltura permettendo anche di monitorare gli effetti che le diverse condizioni di allevamento hanno sulle specie animali.

Gli approcci “*whole-genome*” per lo studio delle interazioni tra i geni e l'ambiente possono permettere il raggiungimento di questo obiettivo; non a caso negli ultimi anni hanno preso piede numerosi programmi di ricerca atti ad aumentare la conoscenza dei genomi/trascrittomi di molte specie allevate. Per i vertebrati superiori, sono state prodotte ampie collezioni di EST (1.560.130 ESTs per il bovino, 2.227.253 per il maiale, 632.013 per il pollo) e sono ora disponibili numerose piattaforme di genomica funzionale, in particolar modo *microarray* a DNA. Per quanto riguarda i teleostei, solo recentemente notevoli sforzi nel sequenziamento hanno permesso di ampliare le collezioni di EST per diverse specie tra cui il salmone (*Salmo salar*; Adzhubei *et al.* 2007), il merluzzo (*Gadus morhua*; Govoroun *et al.* 2006), l'halibut (*Hippoglossus hippoglossus*; Douglas *et al.* 2007) e il pesce gatto (*Ictalurus punctatus*; Li *et al.* 2007). Questa enorme mole di sequenze non può essere analizzata appieno utilizzando la tecnologia *microarray* a cDNA in quanto questa prevede che, per ciascuna sequenza, il clone corrispondente venga spottato nella *slide* e che quindi sia fisicamente disponibile. Questa limitazione determina spesso la costruzione di *microarray* a cDNA in grado di fornire una rappresentazione solo parziale del trascrittoma della specie, riducendo le potenzialità che ne possono derivare (Martin *et al.* 2006; Geoghegan *et al.* 2008; Jørgensen *et al.* 2008). Inoltre, le nuove tecnologie di sequenziamento *high throughput*,

in grado di sequenziare milioni di EST in una singola reazione (Ellegren, 2008; Emrich *et al.* 2007), non usano i cloni batterici come materiale di partenza; l'amplificazione e lo spottaggio di questi ultimi non è quindi più possibile.

Gli *array* a oligonucleotidi forniscono un approccio alternativo alle piattaforme a cDNA, permettendo la rappresentazione di tutte le sequenze disponibili per una determinata specie. Le sonde a oligonucleotidi possono essere di diverse dimensioni (da 24 a 70 nucleotidi) ed essere sintetizzate separatamente e successivamente spottate nella *slide* oppure possono venire sintetizzate *in situ* direttamente nel supporto di vetro.

Fino a poco tempo fa, lo sviluppo di una piattaforma *microarray* a oligonucleotidi richiedeva un investimento economico consistente, come conseguenza del costo della sintesi degli oligo o dello sviluppo di maschere fotolitografiche *ad hoc* (Tecnologia Affymetrix); l'avvento di nuove tecnologie (Nimblegen, Agilent, Combimatrix), che permettono la sintesi *in situ* delle sonde, ha reso abbordabile lo sviluppo di piattaforme *microarray* ad elevata densità anche per le specie “*non-modello*”. Negli ultimi anni infatti, piattaforme a oligonucleotidi sono state progettate per diverse specie di teleostei; un *microarray* a oligo basato su sonde di 24 nucleotidi (24meri) è stato sviluppato per il pesce gatto utilizzando la tecnologia Nimblegen (Li *et al.* 2006; Peatman *et al.* 2008; Liu *et al.* 2008) mentre la Parallel Synthesis Technology è stata applicata per la fabbricazione di una piattaforma ad elevata densità per l'halibut. Le più diffuse sono tuttavia la tecnologia Affimetrix, basata sulla sintesi di sonde *in situ* tramite fotolitografia, e la tecnologia Agilent dove gli oligo vengono sintetizzati mediante una tecnica detta “a getto di inchiostro” (Agilent SurePrint™ technology). La tecnologia Affimetrix è stata ampiamente utilizzata per specie modello come zebrafish (Levi *et al.* 2009) e di recente anche per la trota (Cavileer *et al.* 2009) mentre numerose piattaforme Agilent, diverse sia in formato che in dimensioni delle sonde, sono state progettate per il salmone (Von Schalburg *et al.* 2008), **Pimephales promelas** (Villeneuve *et al.* 2008;

Klaper *et al.* 2008), la sogliola (*Solea senegalensis*; Tingaud-Sequeira *et al.* 2009), il rombo (*Scophthalmus maximus*; Millán *et al.* 2009) nonché per stickleback (Leder *et al.* 2009). In questo studio, tutte le EST a disposizione per *Sparus aurata* sono state impiegate per il *design* di due sonde a oligonucleotidi (60meri) per ciascun trascritto; le sonde sono state sintetizzate *in situ* direttamente in supporti di vetro tramite la tecnologia Agilent *SurePrint*TM. La riproducibilità e l'accuratezza della piattaforma sono state valutate mediante l'analisi dei profili di espressione di due stadi precoci di sviluppo di orata; larve raccolte rispettivamente a uno e quattro giorni dalla schiusa.

3.1.1 La tecnologia Agilent

La tecnologia Agilent produce microarray ad elevata densità tramite la sintesi *in situ* di sonde costituite di 60 nucleotidi (60meri); la sintesi degli oligo si basa sulla tecnologia Agilent *SurePrint*TM definita anche *inkjet* (a getto di inchiostro) in quanto i singoli nucleotidi vengono depositi tramite il getto di micro gocce (del volume di circa 4 pL) di reagente (Hughes *et al.* 2001). La tecnologia Agilent si dimostra essere molto versatile in quanto permette lo sviluppo di piattaforme *custom* a partire da qualunque set di sonde e concede, inoltre, di adattare sia il protocollo di marcatura dei campioni, come anche la scansione e la procedura di analisi delle immagini, a seconda dei diversi disegni sperimentali. I protocolli di preparazione dei campioni possono, infatti, prevedere l'utilizzo di un singolo fluoroforo (Cy-3; approccio *One-color*) o di due (Cy-3 e Cy-5; approccio *Two-color*); nel primo caso i campioni appartenenti alle due diverse categorie vengono entrambi marcati con Cy-3 e ibridati a due array distinti (ibridazione non competitiva); nel secondo caso, invece, gli stessi campioni vengono marcati con due fluorofori distinti (Cy-3 e Cy-5) e successivamente co-ibridati nello stesso array (ibridazione competitiva). Recentemente è stata introdotta anche la possibilità di eseguire esperimenti in *multiplex*, dove diversi set di sonde possono essere sintetizzati e testati indipendentemente nella stessa *slide* (Wolber *et al.* 2006).

3.2 MATERIALI E METODI

3.2.1 Campioni biologici ed estrazione dell'RNA

Esemplari di orata appartenenti a diversi stadi precoci di sviluppo sono stati raccolti presso l'Impianto di Acquacoltura "Ca' Zuliani" (Monfalcone), anestetizzati, congelati interi in azoto liquido e stoccati a -80°C. Per due stadi, lo Stadio 1 (larve raccolte a 24 ore dalla schiusa) e lo Stadio 4 (larve a 96 ore dalla schiusa), l'RNA totale è stato estratto da cinque *pool* indipendenti per stadio utilizzando l' *RNAeasy Mini Kit* (Qiagen, Hilden, Germany); ciascun *pool* era costituito approssimativamente di 40-50 individui. Un ulteriore *pool* è stato invece preparato utilizzando larve appartenenti a quattro diversi stadi di sviluppo; l'RNA estratto da quest'ultimo è stato utilizzato per la preparazione di quattro repliche tecniche allo scopo di testare la riproducibilità, tra *array* indipendenti, della fase di ibridazione.

La qualità dell'RNA è stata preventivamente testata tramite elettroforesi su gel di agarosio all'1% contenente *SYBR Safe™ DNA Gel stain 10,000X* (Invitrogen™, Carlsbad, California) e la concentrazione dell'estratto è stata determinata mediante lettura spettrofotometrica con lo strumento NanoDrop® ND-1000 (NanoDrop Technologies, Wilmington, USA). La qualità e l'integrità dell'RNA è stata inoltre stimata con l'utilizzo del 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA); l'*RNA Integrity Number* (RIN), indice che fornisce una determinazione numerica e quindi oggettiva dell'integrità del campione, è stato, infatti, calcolato per ciascun estratto. Solo campioni dotati di RIN superiore a 7,5 sono stati utilizzati per i successivi esperimenti.

3.2.2 Costruzione del *database* per *Sparus aurata* e disegno delle sonde

Tutte le sequenze di EST di *Sparus aurata* sono state ottenute a partire da 17 librerie normalizzate di cDNA, ciascuna rappresentante un tessuto diverso (fegato, ovario, testicolo, osso/cartilagine, cervello/ghiandola pituitaria, cuore/vasi sanguigni, tessuto

adiposo, rene cefalico, rene, branchie, intestino, milza prelevata da individui sani, milza prelevata in seguito a stimolazione con patogeno, pelle, organo ultimobranchiale, corpuscoli di Stannius).

Il set iniziale di EST a disposizione è stato quindi filtrato per eliminare le sequenze di scarsa qualità e i trascritti rimanenti sono stati esaminati per escludere la presenza delle sequenze del vettore o di elementi ripetuti utilizzando il *software* RepeatMasker.

Le EST raccolte, assieme a tutte le sequenze pubbliche di mRNA di orata, sono state assemblate utilizzando una strategia basata su analisi di Blast per l'identificazione delle sequenze (*e-value* soglia =1E-10) da includere nello stesso *cluster* e successivamente sull'utilizzo del programma Cap3 (Huang *et al.* 1999) per effettuarne l'assemblaggio e produrre le sequenze consenso. Diverse EST erano considerate appartenere allo stesso *cluster* se si sovrapponevano per almeno 40 bp con un'identità della regione sovrapposta di almeno il 90%. L'assemblaggio ha prodotto un set finale di 19.734 diversi *cluster*.

Tutte le EST sono state sottomesse in NCBI, l'*accession number* in GenBank della sequenza che mostra la più elevata identità con ciascun *cluster* è riportata nel "Platform data table" (GPL6467) depositato in GEO (*Gene Expression Omnibus*).

L'annotazione dei trascritti unici ottenuti è stata effettuata sfruttando l'algoritmo di Blast; i criteri di selezione sono stati limitati al miglior *match* con un *e-value* di almeno 1E-10. Questa procedura ha previsto due diversi passaggi:

1. Due ricerche tramite Blastx e Blastn sono state effettuate contro gli interi trascrittomi di *Danio rerio*, *Gasterosteus aculeatus*, *Takifugu rubripes* e *Tetraodon nigroviridis* depositati in ENSEMBL (<http://www.ensembl.org>).
2. Per tutti i geni che non mostravano alcun *match* di similarità nel primo passaggio, una ulteriore ricerca tramite Blastx è stata effettuata contro il database aminoacidico *non redundant* di NCBI.

Le associazioni dei diversi trascritti di orata con i *Gene Ontology* (GO) ID e Pfam ID sono state ottenute, solo per i geni annotati, utilizzando UniprotKB come *database* di riferimento; i *cluster* che mostravano similarità con un'*entry* di Uniprot, infatti, ereditavano automaticamente i suoi GO o Pfam IDs. Allo scopo di avere una visione di insieme del contenuto in GO, i risultati ottenuti dall'annotazione sono stati semplificati utilizzando i termini contenuti nella “*GOA slim*” scaricabili al sito <http://www.geneontology.org/GO.slims.shtml>.

Tutti i trascritti unici di orata e le annotazioni ad essi associate sono stati raccolti nel *database* SAPD (SAPD: *Sparus aurata* Padova), creato sulla base dell'ambiente BioMart. Questo database può essere interrogato utilizzando diversi filtri tra cui l'ID di ciascun *cluster*, la sua descrizione, il GO ID, il Pfam ID o una combinazione di questi criteri; noto l'ID del *cluster* è possibile ottenere diverse informazioni che comprendono la corrispondente sequenza consenso nonché la sua annotazione.

Allo scopo di sviluppare una piattaforma microarray ad elevata densità, per ciascuno dei trascritti unici sono state disegnate due sonde non sovrapposte; il disegno è stato realizzato dal *team* bioinformatico Agilent che ha sfruttato algoritmi di uso esclusivo per questo scopo. La composizione in basi (BC; *Base Composition*) degli oligonucleotidi è stata considerata un indicatore della qualità della sonda; a ciascuna sonda è stato quindi assegnato un punteggio, basato su una scala di cinque livelli (BC1-4, BC *poor*), relativo alle *performance* predette di ibridazione. Il calcolo dei BC si è basato su due metri: la “*Base composition ratio*”, ossia la percentuale delle quattro basi azotate (A, T, G, C) in rapporto tra loro, e le “*Homeomeric runs*”, le ripetizioni di una stessa base presenti all'interno di una sonda e che, riducendone la complessità, aumentano la possibilità di legami aspecifici in fase di ibridazione.

3.2.3 Esperimenti di *microarray* e analisi dei dati

Un totale di 39.379 sonde a oligonucleotidi sono state utilizzate per la costruzione di una piattaforma *microarray* basata sul *design* “Agilent 4x44K” che prevede la sintesi di quattro *microarray*, ciascuno costituito di circa 44.000 sonde, in un’unica *slide*.

La validazione della piattaforma è stata quindi effettuata analizzando i profili di espressione genica di 19.715 trascritti di orata in due stadi precoci di sviluppo: larve a uno e quattro giorni dalla schiusa. La marcatura e l’ibridazione di ciascun campione sono state eseguite in accordo al protocollo “*Agilent One-Color Microarray-Based Gene Expression Analysis*”; brevemente, 500 ng di RNA totale sono stati amplificati e marcati con Cy3-dCTP e una miscela di 10 diversi RNA virali poliadenilati (*Agilent Spike-In Mix*) sono stati inoltre aggiunti a ciascun campione prima dell’amplificazione e marcatura allo scopo di poter monitorare, a posteriori, l’andamento di ciascuna fase dell’esperimento. Il cRNA (RNA complementare) marcato è stato quindi purificato con il kit Qiagen *RNAeasy Mini Kit* e la sua concentrazione, nonché l’attività specifica (pmol Cy3/ μ g cRNA), sono state valutate allo spettrofotometro NanoDrop[®] ND-1000. Un totale di 1.650 ng di cRNA marcato sono stati quindi sottoposti a frammentazione, mediante l’aggiunta di 11 μ l di *Blocking Agent* (10X) e 2.2 μ l di *Fragmentation Buffer* (25X) a cui è seguita un’incubazione a 60°C per 30 min, e diluiti per aggiunta di 55 μ l di *GE Hybridization buffer* (2X). Un volume totale di 100 μ l della soluzione di ibridazione sono stati quindi depositi nel *microarray*, ciascuna *slide* è stata incubata a 65°C per 17 ore e successivamente sottoposta a lavaggio.

Lo scanner Agilent “*G2565BA DNA microarray scanner*” è stato utilizzato per la scansione degli *array* ad una risoluzione di 5 μ m; ciascuna *slide* è stata sottoposta a una duplice scansione utilizzando due diversi livelli di sensibilità del laser (XDR Hi 100% and XDR Lo 10%).

Il software *Feature Extraction 9.5.1* è stato a questo punto utilizzato per l'analisi delle immagini ottenute; le due immagini generate per ciascuna *slide* sono state analizzate contemporaneamente e i dati di espressione sono stati estrapolati e “puliti” dal *background* seguendo la procedura *standard* del programma.

Feature Extraction 9.5.1 restituisce una serie di indicatori della qualità di ciascuno *spot* allo scopo di valutarne l'affidabilità, tra questi, la voce “*glsFound*” (corrispondente a 1 se lo *spot* ha un valore di intensità significativamente superiore al *background*, altrimenti equivalente a 0) è stata utilizzata per filtrare le sonde non attendibili. Per eseguire delle analisi statistiche più robuste ed esenti da fattori di confondimento (*bias*), le sonde con una elevata proporzione di *glsFound* = 0 sono state eliminate dai *dataset*; il numero di *glsFound* = 0 usato come soglia di accettabilità o meno della sonda è stato deciso di volta in volta in base al disegno sperimentale.

Per concludere, le intensità degli *Spike-In* sono state utilizzate per identificare la procedura di normalizzazione migliore per ciascuna serie di dati; dopo normalizzazione, infatti, l'intensità degli *Spike-In* dovrebbe essere uniforme tra i diversi esperimenti del *dataset*. Nel nostro caso, l'approccio *cyclic lowess* (Bolstad *et al.* 2003) si è sempre dimostrato più performante della normalizzazione quantile. La precisione e la ripetibilità della piattaforma *microarray*, sono state valutate calcolando il coefficiente di Pearson sia tra esperimenti che all'interno dello stesso *array* (tra le due sonde di ciascun trascritto). Il filtraggio delle sonde non conformi, le normalizzazioni e le analisi di correlazione sono stati tutti effettuati utilizzando il software statistico R (<http://www.r-project.org>). Per identificare i geni differenzialmente espressi tra lo Stadio 1 e lo Stadio 4 di orata è stato invece utilizzato il test statistico SAM (*Significance Analysis of Microarray*; Thusher *et al.* 2001). Il test non parametrico di correlazione di Spearman è utilizzato per osservare la correlazione tra i valori di espressione misurati rispettivamente con la piattaforma *array* e la metodica di *Real-time RT-PCR*.

3.2.4 Validazione dei dati di espressione mediante *Real-time* RT-PCR

Undici geni *target* sono stati selezionati per l'analisi di espressione mediante *Real-time* RT-PCR; per ciascuno dei geni selezionati e per il gene *reference* (Malate dehydrogenase 1; MDH1) il saggio di espressione è stato progettato utilizzando il sistema di sonde *Universal Probe Library* (UPL) (www.universalprobelibrary.com) (RocheDiagnostic, Mannheim, Germany). I *primer* gene-specifici, nonché la sonda UPL corrispondente, sono stati determinati col *software* ProbeFinder (vedi Appendice 3A).

Per poter disegnare delle sonde a cavallo di un introne, le giunzioni esone-introne dei trascritti di orata, ove possibile, sono state dedotte sulla base del confronto con le sequenze genomiche di omologhi in specie di teleostei per i quali i genomi sono stati interamente sequenziati e resi pubblici (*T. nigroviridis*, *D. rerio* e *G. aculeatus*).

I campioni testati in *Real-time* RT-PCR sono gli stessi utilizzati negli esperimenti di *microarray*, incluso il calibratore, controllo interno per ciascuna reazione di amplificazione. Un microgrammo di RNA totale di ciascun campione è stato retrotrascritto a cDNA utilizzando l'enzima Superscript II (Invitrogen™); una aliquota (2.5 µl) di cDNA diluito 1:40 è stato quindi amplificato in un volume totale di 10 µl contenente 5 µl di *FastStart TaqMan® Probe Master 2X* (Roche Diagnostics), 0,25 µl di ciascun *primer* gene-specifico (10 µM) e 0,1 µl di sonda UPL (100 µM). Tutti gli esperimenti sono stati effettuati con lo strumento LightCycler® 480 (Roche Diagnostics); il protocollo di amplificazione ha previsto uno *step* iniziale di 2 min a 50°C e 10 min a 95°C seguito da 45 cicli di 10'' a 95°C e 30'' a 60°C.

Per valutare l'efficienza di ciascun saggio, curve standard sono state costruite amplificando diluizioni seriali (1:2) dello stesso cDNA utilizzato come calibratore (Sa_stage1_A). Il valore di Cp (*Crossing point*) è stato utilizzato per determinare la concentrazione relativa del gene *target* in ciascun campione; tutte le quantificazioni sono state effettuate in duplicato e normalizzate al gene *reference*.

3.3 RISULTATI

3.3.1 Costruzione del *database* SAPD

Un totale di 59.485 ESTs e 157 sequenze di mRNA disponibili in GenBank sono state sottoposte ad assemblaggio allo scopo di ottenere sequenze di trascritti unici di orata. In questo modo è stato ottenuto un totale di 19.734 *cluster*; tra questi, 3.563 sono contigui, ossia sequenze formate da due o più EST, mentre le restanti 16.171 sono sequenze singole (*singleton*). Un numero così rilevante di *singleton* è probabilmente dovuto a due fattori: in primo luogo i criteri molto stringenti applicati nella fase di assemblaggio (Vedi Paragrafo 3.2.2) allo scopo di evitare la formazione di concatameri; ed, in seconda istanza, dal fatto che la maggior parte delle EST derivano da librerie di cDNA normalizzate, cosa che ha determinato un basso numero di contigui rispetto ai *singleton*. Il processo di annotazione ha permesso di identificare 8.021 *cluster* che mostrassero una significativa similarità con un gene noto e 7.913 trascritti potenzialmente codificanti una proteina con uno o più domini Pfam. A una buona porzione di contigui, inoltre, è stato possibile associare un GO ID alternativamente per “*Biological process*” (3.332 *cluster*) o “*Cellular component*” (2.301 *cluster*) o “*Molecular function*” (4.420 *cluster*). I termini GO più rappresentati per ciascuna categoria sono raffigurati in Appendice 3B. Tutte le 19.734 sequenze di *Sparus aurata* e le annotazioni ad esse associate sono state depositate in un *database* dedicato e di pubblico accesso: *SAPD database* (<http://enne.cribi.unipd.it:5555/biomart/martview>). Questo *database* è organizzato sulla base dell’ambiente Biomart, il quale permette diverse opzioni per il recupero e l’utilizzo delle informazioni in esso contenute. L’opzione di ricerca tramite similarità di sequenza, per mezzo dell’algoritmo di Blast, è inoltre implementata nel sito.

3.3.2 Analisi dei dati di *microarray*

3.3.2.1 Valutazione delle performance della piattaforma *microarray*

Il disegno degli oligonucleotidi è stato possibile per 19.715 trascritti *target*; di questi, 19.664 sono rappresentati da due sonde non sovrapposte mentre per 51 sequenze è stato possibile disegnare solo una sonda. Un totale di 39.379 sonde *target* sono state quindi sintetizzate direttamente nel supporto di vetro; alla maggior parte di esse (96,1%) è stato attribuito il livello massimo di qualità (BC1), al 3,6% è stato assegnato il livello BC2 e al rimanente 0,3% i punteggi BC3 o BC4; nessuna sonda invece è stata indicata essere di scarsa qualità (*BC poor*).

La qualità di ciascuna sonda, predetta *in silico*, è stata quindi testata sperimentalmente in base al numero di ibridazioni riuscite su un totale di 10 esperimenti (cinque repliche biologiche per ciascuno dei due stadi di *S. aurata* testati); le ibridazioni sono state considerate “riuscite” solo per le sonde alle quali era associato un *glsFound* = 1 dal programma *Feature Extraction 9.5.1*. Solo cinque sonde (0,013%) non hanno mai mostrato un’intensità di segnale superiore al rumore di fondo mentre per 37.585 oligonucleotidi, corrispondenti al 95% del numero totale di sonde, le ibridazioni sono riuscite in almeno cinque esperimenti su dieci (Figura 3.1).

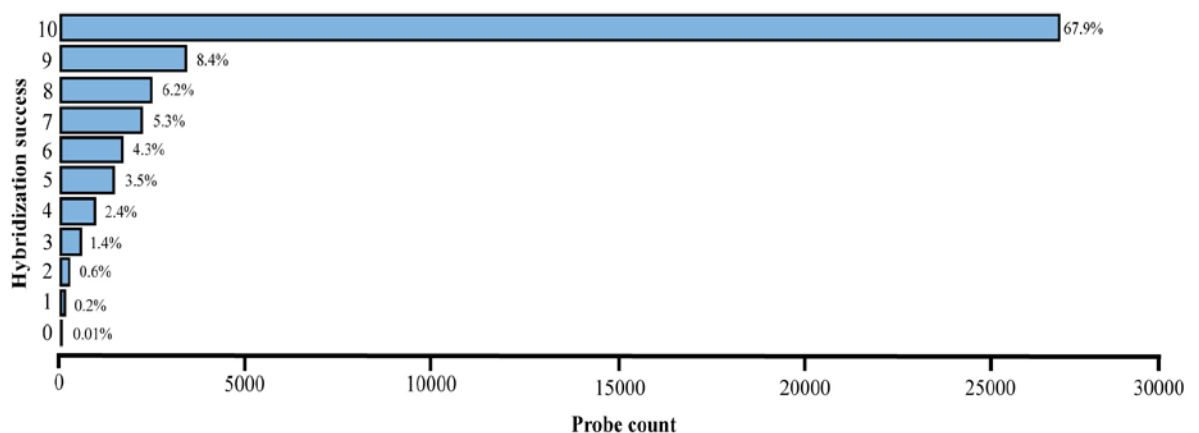


Figura 3.1: Analisi della distribuzione del successo di ibridazione di ciascuna sonda. Per ciascuna sonda è stato calcolato il numero di ibridazioni riuscite (*glsFound* = 1) Asse Y: numero di ibridazioni riuscite in 10 esperimenti (0 corrisponde al numero di sonde che non hanno mai ibridato, 1 a quelle per le quali le ibridazioni sono sempre riuscite. Asse X: numero di sonde che ricade in ciascun gruppo (0-10).

Uno dei requisiti più importanti di un esperimento di microarray è la riproducibilità dell'intero sistema, che assicura che i risultati ottenuti da esperimenti indipendenti possano essere direttamente comparabili. Allo scopo di valutare la riproducibilità e la precisione del protocollo sperimentale e della piattaforma stessa, sono state effettuate quattro repliche tecniche dello stesso campione. I dati grezzi di fluorescenza sono stati filtrati in base al numero di ibridazioni non riuscite ($glsFound = 0$) per ciascuna sonda; considerate tutte e quattro le repliche, gli oligo con più di un $glsFound = 0$ sono stati rimossi dalle analisi successive (3.224 sonde rimosse, 8% del numero totale di sonde). In seguito alla normalizzazione *cyclic lowess*, il grado di concordanza tra le repliche è stato calcolato tramite la stima del coefficiente di correlazione di Pearson per l'intero set di valori di fluorescenza. Per ciascuna coppia di esperimenti, i coefficienti di correlazione sono risultati essere sempre significativi ($p\text{-value} < 1E-5$) e mai inferiori allo 0,99 dimostrando, quindi, una correlazione quasi perfetta tra le repliche ed una buonissima ripetibilità della piattaforma stessa.

In aggiunta, è stato calcolato anche il coefficiente di variazione percentuale (CV%) dei valori di fluorescenza normalizzati tra le sonde *target* nelle quattro repliche sperimentali; il valore mediano del CV% è risultato essere 1,1% e meno del 15% delle sonde hanno presentato un CV% superiore al 30% (2.562 sonde su 37.592).

Come per le repliche tecniche, i dati di fluorescenza grezzi, ottenuti dal confronto dei due stadi di sviluppo di *S. aurata*, sono stati filtrati in base al successo di ibridazione di ciascuna sonda; in questo caso, gli oligo con più di due $glsFound = 0$ tra la cinque repliche biologiche di ciascuno stadio sono stati rimossi dalle analisi successive (4.846 sonde rimosse, 12% del numero totale di sonde).

La piattaforma microarray di orata è caratterizzata dalla presenza di due sonde per ciascun trascritto; queste due sonde si appaiano alla sequenza genica in due regioni non sovrapposte, in particolare, la prima sonda (Sonda_1) è progettata per essere il più vicino possibile al 3' di ciascun trascritto. Alla luce di questo, la variabilità tra le due sonde per ciascun trascritto è stata valutata utilizzando la differenza dei due valori di intensità come misura della difformità di segnale. Questo confronto dovrebbe dare una differenza il più vicina possibile (se non uguale) allo zero; con l'eccezione di pochi casi, infatti, la maggior parte delle coppie di sonde sono caratterizzate da una piccola differenza in termini di intensità di fluorescenza.

In Figura 3.2 (A.), ciascun *boxplot* mostra la distribuzione delle differenze di segnale tra Sonda_1 e Sonda_2 in ciascun esperimento; come atteso, in tutti gli esperimenti le differenze tra le intensità delle due sonde per lo stesso gene mostrano una distribuzione simmetrica e centrata sullo zero.

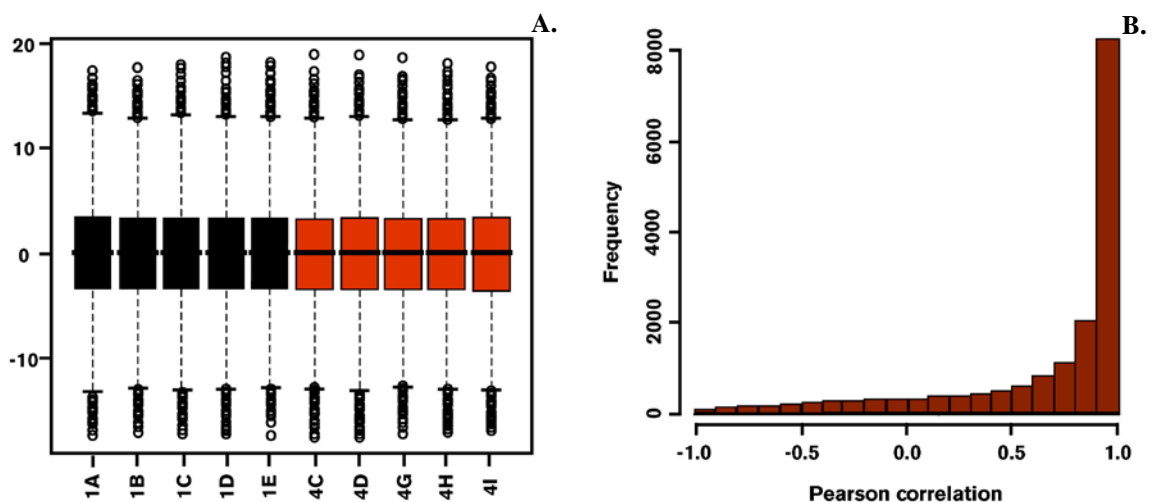


Figura 3.2: A. Boxplot della distribuzione delle differenze di intensità osservate (asse y), tra le due sonde (Sonda_1 e Sonda_2) di ciascun trascritto. Le sigle nell'asse x si riferisce ai seguenti esperimenti: 1A-1E, repliche biologiche dello Stadio 1; 4C-4I, repliche biologiche dello Stadio 4. B. Distribuzione della correlazione tra i valori di espressione misurati dalla Sonda_1 e Sonda_2. Per ciascun gene, la correlazione di Pearson è stata stimata tra esperimenti e all'interno dello stesso.

Allo scopo di valutare il grado di concordanza fra i valori di espressione delle coppie di sonde per ciascun trascritto, il coefficiente di correlazione di Pearson è stato calcolato sia tra esperimenti che all'interno dello stesso esperimento; in quest'ultimo caso i valori di espressione tra Sonda_1 e Sonda_2 mostrano dei coefficienti di correlazione sempre superiori allo 0,8. La correlazione tra *microarray* è stata invece calcolata utilizzando i vettori dei valori di espressione, rispettivamente della Sonda_1 e della Sonda_2, tra tutti e 10 gli esperimenti; la distribuzione dei coefficienti di correlazione osservati (Figura 3.2 B.) indica che la maggior parte delle sonde (68%) hanno una buona correlazione positiva ($r > 0,7$), l'11% una correlazione moderata ($0,5 < r < 0,7$), mentre solo una piccola parte di sonde (3%) sono correlate negativamente.

Un altro requisito importante per qualunque piattaforma *microarray* è che questa copra un ampio *range* dinamico, in modo da poter rilevare e quantificare sia i trascritti più rari che quelli maggiormente espressi all'interno dello stesso esperimento. La sensibilità e il *range* dinamico della piattaforma sono stati entrambi misurati utilizzando come riferimento le sonde di controllo *Spike-In*. Gli *Spike-In* sono costituiti da una miscela di 10 trascritti poliadenilati e sintetizzati *in vitro* derivanti dal gene E1A dell'Adenovirus; le concentrazioni di questi RNA coprono un *range* di sei ordini di grandezza (da 0,04 pg/μl a 40.000 pg/μl). Interpolando il segnale di fluorescenza per ciascuno dei trascritti *Spike-In* al valore logaritmico della concentrazione corrispondente, il *range* di quantificazione lineare può essere calcolato in base alla retta di interpolazione dei dati. Il limite di rilevazione (LLD; *Lower Limit of Detection*) degli esperimenti è stato calcolato considerando la più bassa concentrazione di *Spike-In* all'interno del *range* lineare; in tutti gli esperimenti considerati è stato osservato un *range* dinamico molto ampio con un incremento lineare dell'intensità di segnale lungo cinque ordini di grandezza (4.96 ± 0.2) ed un LLD di 0,4 pg/μl (corrispondente allo *Spike-in* E1A_r60_a104).

3.3.2.2 Analisi dei geni differenzialmente espressi

Allo scopo di identificare i geni differenzialmente espressi tra gli stadi di sviluppo di orata Stadio 1 e Stadio 4, è stata eseguita un'analisi statistica mediante l'utilizzo di SAM (SAM a due classi) con un *False Discovery Rate* (FDR) uguale a zero.

Questo ha prodotto una lista di 1.518 (4%) sonde significative corrispondenti a 1.050 geni; per 468 di questi, sia la Sonda_1 che la Sonda_2 sono risultate essere differenzialmente espresse mentre per i rimanenti 582 geni solo una delle due sonde è stata indicata come significativa dall'analisi di SAM.

Un totale di 643 trascritti (di cui 289 rappresentati da entrambe le sonde) sono stati dimostrati essere sovraespressi nello Stadio 4 rispetto allo Stadio 1 mentre il numero di geni risultati sottoespressi è 407 (179 con entrambe le sonde); di questi, una prima annotazione è stata possibile solo per 133 (21%) dei trascritti sovraespressi mentre una percentuale significativamente maggiore (Test di Fisher $p < 0.0001$) dei geni sottoespressi aveva una descrizione associata (283 geni, corrispondenti al 70%).

A 134 (33%) dei trascritti sottoespressi è stato inoltre possibile associare un termine GO per "*Biological process*"; tra questi geni, 38 sono risultati essere coinvolti nella replicazione o riparazione del DNA, assemblaggio della cromatina e regolazione del ciclo cellulare, mentre 36 prendono parte ai processi di sintesi/maturazione delle proteine o catabolismo delle stesse; il terzo gruppo più rappresentato è risultato essere il metabolismo e trasporto lipidico. Per contro, solo 52 (8%) geni sovraespressi avevano un GO ID per "*Biological process*" associato; tra questi, il gruppo più numeroso (proteolisi, 9 termini) è costituito da proteasi con diverse funzioni, come per esempio enzimi digestivi (*chymotrypsinogen*, *elastase*) o peptidasi degli antigeni (*cathepsin L1*). Trasduzione del segnale è il secondo processo biologico più frequente (7 termini) mentre è da sottolineare la presenza di due proteine coinvolte nella fototrasduzione (*retinal cone arrestin-3* e *green-sensitive opsin-1*) e del recettore nucleare dei

glucocorticoidi (*Nuclear receptor 3 C1*). Altre categorie GO per “*Biological process*”, rappresentate da un numero ridotto di termini, sono “processo metabolico”, che in gran parte consiste di enzimi per il metabolismo dei carboidrati, e “trasporto” con trasportatori/canali per diverse molecole (ioni, lipidi).

I dati di fluorescenza, sia grezzi che normalizzati, sono stati depositati nel *database* GEO con *accession number* GSE12116 e GSE12118.

3.3.3 Analisi di espressione tramite *Real-time* RT-PCR

Allo scopo di validare le *performance* della piattaforma *microarray*, saggi di espressione in *Real-time* RT-PCR sono stati effettuati utilizzando il sistema *Universal Probe Library* (UPL). I geni *target* per le analisi di PCR quantitativa (qPCR) sono stati selezionati sulla base dei seguenti criteri: i) dovevano essere rappresentativi dell'intero *range* di *fold-change* riscontrato (1,25-44 volte); ii) sia i geni sottoespressi che quelli sovraespressi dovevano essere equamente rappresentati; iii) per tutti i geni *target*, entrambe le sonde dovevano essere presenti nella serie dei dati normalizzati.

Tabella 3.1: Confronto dei valori di *fold-change* ottenuti rispettivamente con *Real-time* RT-PCR, Sonda_1 e Sonda_2 della piattaforma *microarray* per *S. aurata*

TRASCritto TARGET	SAPD ID	<i>Fold-change</i> ^a		
		<i>Real-time</i> RT-PCR	Microarray Sonda 1	Microarray Sonda 2
Apolipoprotein E1	SAPD02358	0.00362	0.12263	0.1276
Flap endonuclease 1	SAPD04884	0.07811	0.16678	0.18717
Ovostatin	SAPD01680	0.00525	0.13731	0.14149
Myosin	SAPD10294	0.16533	0.25064	0.21589
Serotransferrin	SAPD01126	0.02439	0.11883	0.10408
Glutamate R7	SAPD19202	3.58394	4.30343	4.61982
Methionine aminopeptidase 2	SAPD26496	12.9384	6.88419	6.91868
L-lactate dehydrogenase	SAPD00597	3.09142	2.99927	3.00815
Serine racemase	SAPD19150	40.9397	18.3846	19.9929
Retinal cone arrestin-3	SAPD02277	2232.6	29.4717	44.2798
Phosphoglycerate kinase 1	SAPD03464	1.84119	1.26289	1.12021
Malate dehydrogenase 1	SAPD02236		1.05	0.99

^a Il *fold-change* è stato calcolato come rapporto Stadio 4 / Stadio 1, utilizzando la media dei valori di intensità delle cinque repliche biologiche di ciascuno stadio. Valori <1 indicano sottoespressione nello Stadio 4.

In Tabella 3.1 sono rappresentati i *fold-change* rilevati, per il medesimo trascritto, rispettivamente dal saggio in *Real-time* RT-PCR e da entrambe le sonde (Sonda_1 e Sonda_2) del *microarray*.

Per tutti i geni *target*, la direzione del cambiamento di espressione è uniforme tra qPCR e i dati di *microarray*; una buona concordanza tra i risultati di entrambe le metodiche è stata osservata per valori di *fold-change* compresi tra 2 e 7, quando si superano le 10 volte, invece, i valori di *fold-change* apprezzati dalla *Real-time* RT-PCR sono decisamente superiori a quelli stimati dalla piattaforma *microarray* (Figura 3.3)

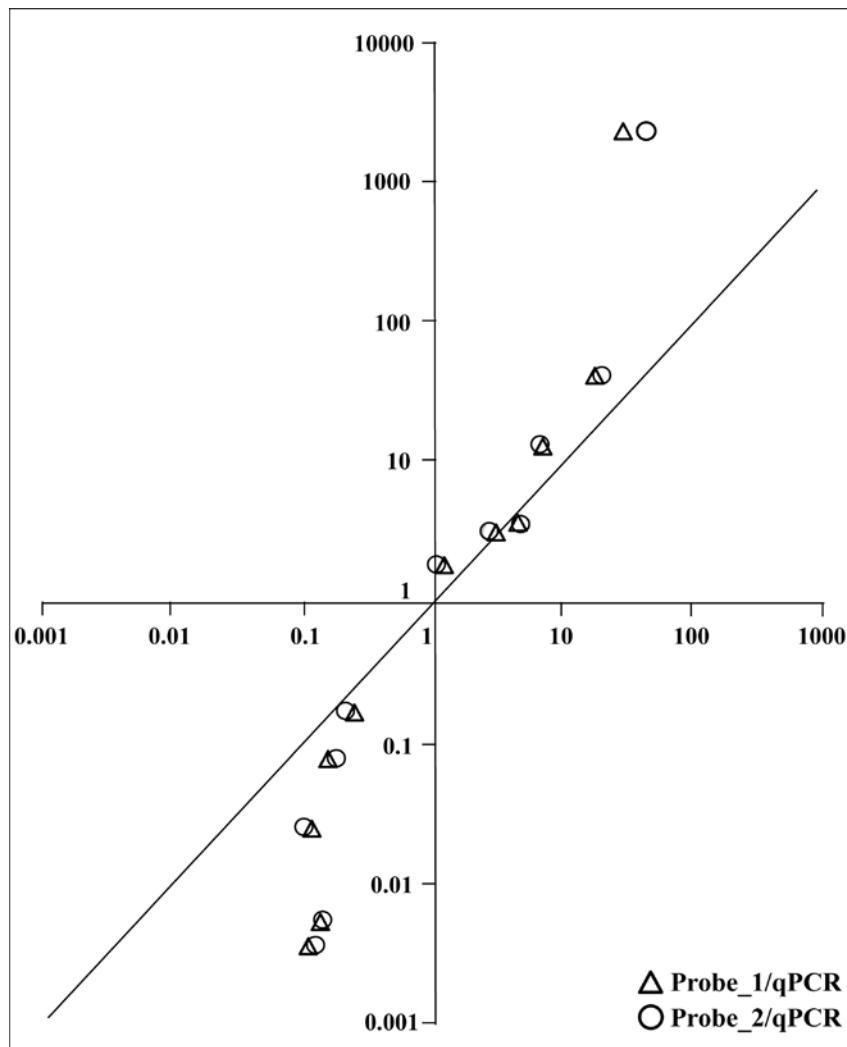


Figura 3.3: Confronto tra i *fold-change* calcolati dalla piattaforma *microarray* e la qPCR. *Triangoli:* rapporto tra i *fold-change* stimati da Sonda_1 (Probe_1) e qPCR *Cerchi:* rapporto tra i *fold-change* stimati da Sonda_2 (Probe_2) e qPCR.

Complessivamente, una correlazione statisticamente significativa tra qPCR e *microarray* è stata ottenuta confrontando i valori di espressione (media dei valori ottenuti tra tutte le repliche biologiche di ciascuno stadio) di ciascun gene *target* (Tabella 3.2). Sei geni hanno mostrato elevati coefficienti di correlazione (Spearman $\rho > 0,8$) per entrambe le sonde ($p < 0,01$) e altri quattro trascritti hanno riportato una significativa correlazione positiva ($0,6 < \rho < 0,8$ con $p < 0,05$); solo un gene, PGK1 (*Phosphoglycerate kinase 1*), ha mostrato una correlazione positiva ma non statisticamente significativa ($\rho = 0,5$, $p > 0,1$) per la Sonda_1 mentre assenza di correlazione è stata osservata per la Sonda_2 ($\rho = -0,04$).

Tabella 3.2: Correlazione tra i valori di espressione rilevati dalla *Real-time* RT-PCR e le due sonde *microarray* per ciascun trascritto.

Gene	Rho di Spearman qPCR/Sonda_1	Rho di Spearman qPCR/Sonda_2	Rho di Spearman Sonda_1/Sonda_2
Apolipoprotein E1	0.915**	0.915**	0.976**
L-lactate dehydrogenase	0.697*	0.636*	0.927**
Methionine aminopeptidase 2	0.939**	0.709*	0.806**
Myosin	0.644*	0.767**	0.867**
Retinal cone arrestin-3	0.905**	0.851**	0.952**
Ovostatin	0.855**	0.855**	0.855**
Phosphoglycerate kinase 1	0.515	-0.042	0.685*
Serine racemase	0.950**	0.917**	0.988**
Serotransferrin	0.732*	0.794**	0.891**
Flap endonuclease 1	0.723*	0.608	0.903**
Glutamate R7	0.818**	0.915**	0.915**

* $p < 0,05$ ** $p < 0,001$

3.4 DISCUSSIONE

Lo scopo di questo lavoro è stato lo sviluppo di una piattaforma *microarray* per lo studio dei profili di espressione genica in orata. La prima fase ha previsto la costruzione di un database di trascritti unici di *S. aurata* assemblando tutte le sequenze pubbliche di mRNA e più di 50.000 EST derivanti da 17 diverse librerie normalizzate di cDNA.

Il numero di trascritti unici ottenuti è simile a quanto riportato per collezioni di EST comparabili, costruite per altre specie/stadi di pesci (stickleback, medaka, pesce gatto, salmone, halibut ecc); circa il 40% di questi, infatti, ha trovato una *match* significativo con almeno un gene o proteina depositati in *database* pubblici (vedi Paragrafo 3.2.2), percentuale solo leggermente inferiore a quella osservata nel salmone (50,3%, Adzhubei *et al.* 2007), l'halibut (60%, Douglas *et al.* 2007) e il pesce gatto (51%, Liu *et al.* 2008).

Il numero relativamente basso di sequenze annotate è una delle limitazioni principali di molti progetti di sequenziamento di EST in specie non-modello, anche in quelle per le quali il trascrittoma è stato caratterizzato a fondo; ci si aspetta, tuttavia, che la percentuale di trascritti annotati possa aumentare sensibilmente nel prossimo futuro, quando, cioè, i genomi completi di nuove specie di teleostei (tilapia, salmone) saranno disponibili. Nuove sequenze e quindi nuove informazioni applicabili nella genomica comparativa, avranno origine anche dall'applicazione delle nuove tecnologie di sequenziamento *ultra-high throughput*.

Il numero relativamente basso di EST disponibili per l'orata non ha inficiato l'efficienza del *design* delle sonde per la piattaforma *microarray*; per la maggior parte dei *cluster*, infatti, è stato possibile disegnare due sonde non sovrapposte. Una elevata correlazione è stata inoltre osservata per la maggior parte delle coppie di sonde disegnate per ciascun trascritto. Solo per 385 trascritti (3%), la Sonda_1 e la Sonda_2 hanno mostrato una correlazione negativa; questo può essere spiegato da diversi fattori: i) *splicing* alternativi possono dare origine a trascritti differenzialmente espressi per lo stesso gene

di origine, questa differenza può quindi essere apprezzata utilizzando due sonde diverse per gene; ii) una maggiore stabilità dell'estremità 3' di certi trascritti può ridurre il segnale della sonda localizzata al 5'; questo, tuttavia, non sembra essere un fenomeno generalizzato in quanto non è stato osservato nessun *bias* significativo tra le sonde al 3' e quelle al 5'; iii) una elevata similarità di sequenza tra geni diversi (per es. loci duplicatisi di recente) può dare origine al problema, ampiamente documentato, della cross-ibridazione o della creazione di *cluster* spuri come conseguenza di errori di assemblaggio.

Prima della normalizzazione e dell'analisi statistica, i dati di circa il 12% delle sonde sono stati rimossi utilizzando un criterio molto stringente (per ciascuna sonda è stato concesso un massimo di due spot "mancanti" tra le cinque repliche biologiche); questa fase di filtraggio è stata eseguita per massimizzare la probabilità di identificare differenze reali nell'espressione genica a scapito della perdita di alcune informazioni. Un'analisi dettagliata delle sonde eliminate ha mostrato che il 60% delle sonde escluse allo Stadio 1 venivano rilevate allo Stadio 1 e, viceversa, il 65% degli spot risultati "mancanti" allo Stadio 4 erano invece presenti allo Stadio 1; questa osservazione suggerisce che, in realtà, la causa dell'eliminazione di un numero relativamente elevato di sonde potrebbe essere una differenza di espressione ontogenetica piuttosto che la scarsa qualità delle sonde stesse. È importante notare, inoltre, che per meno del 4% (769) dei geni rappresentati nell'*array* sono state eliminate entrambe le sonde; per la maggior parte dei trascritti una (3.308 geni) o due sonde (15.638 geni) sono state incluse nelle analisi statistiche; questo chiaramente suggerisce che l'utilizzo di due sonde per trascritto rappresenta un approccio "tutelante" nel *design* della piattaforma microarray.

Una buona ripetibilità della piattaforma Agilent e di altri *array* a oligonucleotidi era già stata documentata da un'ampia iniziativa di controllo della qualità dei *microarray*, operata dal consorzio MAQC (*MicroArray Quality Control*; MAQC Consortium, 2006), anche nel caso del *microarray* di orata la ripetibilità dei dati, sia tra repliche tecniche che biologiche, è risultata elevata e non appare essere influenzata dalla conoscenza, attualmente ancora limitata, del trascrittoma di questa specie. All'interno dell'iniziativa MAQC, sono state confrontate anche le *performance* degli approcci a uno o due colori (Patterson *et al.* 2006); questa comparazione ha stabilito che la qualità dei dati ottenuti con i due approcci di marcatura è sostanzialmente equivalente, indicando che questo fattore non dev'essere di primaria importanza nella scelta del disegno sperimentale. L'uso della sola Cy-3 permette disegni sperimentali più semplici, i rischi di degradazione del fluoroforo ad opera dell'ozono sono ridotti (la Cy-5 è molto più sensibile) e generalmente è meno dispendioso; tuttavia, l'approccio a un colore richiede una normalizzazione tra gli esperimenti molto efficace. Sulla base del confronto del segnale degli *Spike-In* dopo la normalizzazione, l'approccio *cyclic lowess* è risultato essere più performante rispetto alla normalizzazione quantile e decisamente migliore se confrontato al metodo di normalizzazione, basato sulla divisione di tutti i valori di espressione per il valore mediano di fluorescenza, suggerito da Agilent per gli esperimenti ad un colore.

La qualità del *microarray* a DNA progettato per *S. aurata* è stata ulteriormente confermata dalla validazione dei dati di espressione, di un *set* di geni *target* selezionati, mediante *Real-time* RT-PCR. L'utilizzo della qPCR per la cross-validazione dei risultati di *microarray* generalmente è limitata ai geni che più significativamente risultano essere differenzialmente espressi; in questo studio, invece, i geni scelti per la validazione ricoprono l'intero *range* di intensità di segnale nonché di *fold-change* osservati. Sebbene questo approccio non possa sostituire un'analisi sistematica mediante qPCR di

tutti i geni rappresentati nell'*array*, riportata invece in altri studi (Wang *et al.* 2006), dovrebbe in ogni caso permettere un confronto dei risultati ottenuti con le due diverse metodiche meno influenzato da *bias*. Nel caso dell'orata, è stata riscontrata una buona correlazione positiva tra le due tecnologie, confermando ulteriormente l'affidabilità della piattaforma *microarray*; l'unica eccezione è rappresentata dal gene PGK1 per il quale una correlazione moderatamente positiva ma non statisticamente significativa è stata osservata solo tra Sonda_1 e qPCR. Questo probabilmente è da attribuirsi alla differenza di espressione molto piccola, per questo gene, tra i due gruppi di campioni (*fold-change* medio stimato dal *microarray* = 1,1-1,25 volte); la scarsa correlazione tra *microarray* e *Real-time* RT-PCR per i geni che esibiscono piccole differenze di espressione (< 1,4 volte) è, infatti, comunemente riportato, tant'è che un *fold-change* di due è considerato la soglia al di sotto della quale le due tecnologie perdono corrispondenza (Morey *et al.* 2006). Rapportando i valori di *fold-change* stimati dal *microarray* con i corrispondenti quantificati dalla qPCR (Figura 3.3) è possibile, inoltre, osservare l'evento noto col nome di "*fold-change compression*" per differenze di espressione che superano un ordine di grandezza; questo è un fenomeno ampiamente riportato e dovuto a diverse limitazioni tecniche come, per esempio, un range dinamico inferiore rispetto alla qPCR, saturazione del segnale e fenomeni di cross-ibridazione nei *microarray* (Wang *et al.* 2006).

Nonostante l'obiettivo principale di questo lavoro fosse lo sviluppo e la validazione di una piattaforma *microarray* per lo studio dei profili di espressione genica in *S. aurata*, è stato anche possibile ottenere dei risultati significativi riguardo i processi biologici in atto nelle fasi precoci di sviluppo dell'orata. E' tuttavia da sottolineare che, poiché gli esperimenti hanno avuto come oggetto larve intere, i livelli di espressione di ciascun gene *target*, rappresentato nell'*array*, riflettono un ampio insieme di tessuti e tipi cellulari diversi; per questo motivo le differenze di espressione tra Stadio 1 e Stadio 4

potrebbero essere dovute al cambiamento nelle proporzioni dei diversi tessuti piuttosto che a modificazioni reali dei livelli di espressione dei singoli geni *target*.

Allo stesso modo, l'assenza di differenze di espressione potrebbe essere dovuta all'annullamento delle stesse a causa di profili di espressione opposti, per lo stesso gene, in tessuti diversi.

I trascritti che risultano essere sottoespressi nello Stadio 4 rispetto allo Stadio 1 appartengono principalmente alla categoria “geni *housekeeping*”, ossia geni coinvolti in processi biologici essenziali (come la replicazione del DNA, il ciclo cellulare e la sintesi proteica); è quindi probabile che, man mano che il differenziamento cellulare o tissutale procedono, comincino ad essere prodotti trascritti più specifici, determinando una “diluizione” dei trascritti codificanti proteine *housekeeping*. Un motivo simile potrebbe essere la causa della sottoespressione riscontrata per le proteine coinvolte nel metabolismo lipidico, essenziale per la sintesi delle membrane cellulari; d'altra parte però, nello Stadio 4 il sacco vitellino è ridotto a 1/9 della sua dimensione originale con una corrispondente riduzione del contributo dei lipidi, in esso contenuti, come nutrienti; è quindi possibile che la sottoespressione di mRNA codificanti proteine coinvolte nel metabolismo lipidico rifletta la transizione verso l'alimentazione autonoma delle larve.

Nelle larve al quarto giorno dalla schiusa, inoltre, il sistema digestivo è formato e la bocca inizia ad aprirsi; in accordo con questi eventi, enzimi digestivi come l'elastasi, già riportata da Sarropoulou e colleghi (2005), nonché due isoforme di chimotripsinogeno appaiono sovraespressi; come appaiono sovraespressi alcuni geni specifici dell'occhio (*green-sensitive opsin*, *retinal cone arrestin-3* e *cathepsin L2*, coinvolta nello sviluppo della cornea), in corrispondenza della formazione di occhi pigmentati, tipica di questa fase di sviluppo larvale.

Un altro processo ben avviato durante lo sviluppo larvale è la miogenesi; la differenziazione delle fibre muscolari coinvolge una complessa serie temporale di

attivazioni geniche (Xu *et al.* 2000; Hall *et al.* 2003; Chauvigné *et al.* 2005) che coinvolge proteine strutturali e contrattili (per es. *myosin*, *tropomyosin*) come anche proteine muscolari solubili ed altri enzimi (per es. *parvalbumin*, *muscle creatine kinase*); sfortunatamente sono disponibili ancora poche informazioni a proposito dell'organizzazione spaziale e temporale dell'espressione genica nei pesci durante la miogenesi.

In questo studio, elevati livelli di espressione del fattore MyoD sono stati riscontrati sia nello Stadio 1 che nello Stadio 4; similmente, sono risultati essere abbondantemente espressi trascritti coinvolti nella contrazione muscolare, come *myosin light chain 1*, *parvalbumin*, *tropomyosin*, e *sarcomeric creatine kinase* (CKM). Quest'ultima mostra una forte sovraespressione nello Stadio 4, confermando le precedenti evidenze di un costante incremento dei livelli di espressione di CKM dallo stadio embrionale a quello adulto (Xu *et al.* 2000). Come già riportato da Sarropoulou e colleghi (2005), differenze nei livelli di espressione sono state osservate anche per la tropomiosina, per la quale i livelli di espressione aumentano con la crescita dell'embrione, mentre sia la miosina che la parvalbumina mostrano una debole sovraespressione (inferiore alle 4 volte) nello Stadio 1 rispetto allo Stadio 4; per concludere, lo *stromal cell derived-factor*, molecola che promuove la differenziazione miogenica delle cellule precursori (Rescan *et al.* 2008), appare sottoespressa nello Stadio 4 rispetto allo Stadio 1.

Lo spostamento dai trascritti "essenziali" a quelli tessuto-specifici può fornire una spiegazione alla significativa differenza nella percentuale di geni annotati tra i trascritti sovra e sottoespressi nello Stadio 4; è stato osservato, infatti, un numero piuttosto basso di *cluster* annotati tra i trascritti sovraespressi (annotati per il 21%) rispetto a quelli sottoespressi (annotati per l'80%). Essendo l'annotazione dei *cluster* basata essenzialmente sulla similarità di sequenza; è quindi più probabile che trascritti di orata derivanti da geni altamente conservati trovino *match* significativi con sequenze di altre

specie; è stata infatti riportata in diversi studi (Duret *et al.* 2000; Lehner *et al.* 2004; Subramaniam *et al.* 2004; Zhang *et al.* 2004) una correlazione tra il livello di conservazione della sequenza e la funzione/distribuzione/espressione della proteina da essa codificata e sembra che, almeno tra i mammiferi, geni “essenziali” o *housekeeping* (espressi ubiquitariamente) evolvano più lentamente rispetto a proteine non-essenziali o tessuto-specifiche.

Nel caso dell’orata, la transizione da Stadio 1 a Stadio 4 determina un aumento dei trascritti tessuto- o cellulo-specifici e questo probabilmente si traduce in un contenuto maggiore di geni *housekeeping* nello Stadio 1; poiché un numero significativamente più elevato di geni sottoespressi allo Stadio 4 mostra similarità di sequenza con potenziali geni omologhi in altre specie, sembra probabile che anche in orata i geni essenziali e/o *housekeeping* evolvano meno velocemente.

Processi di selezione simili sembrano, quindi, direzionare l’evoluzione dei geni codificanti per proteine sia nei vertebrati inferiori che in quelli superiori.



**4. SVILUPPO DI UNA
PIATTAFORMA *MICROARRAY* PER
LO STUDIO DEL PROGNATISMO
IN *DICENTRARCHUS LABRAX***

4.1 INTRODUZIONE

Il prognatismo mandibolare (MP; *Mandibular Prognathism*) o protrusione del labbro inferiore, è una malformazione dello sviluppo che determina un fenotipo facciale particolarmente distintivo. Uno degli esempi più noti di prognatismo è la famiglia reale degli Asburgo, dove questa malformazione è stata riscontrata in diverse generazioni successive (Chudley, 1998) suggerendo una forte componente genetica come causa di questo disordine; infatti, anche se fattori ambientali sembrano contribuire allo sviluppo del MP, è stata riportata aggregazione familiare in diverse popolazioni umane, supportando l'ipotesi che componenti ereditarie giochino un ruolo nell'eziologia di questa malformazione. Un recente studio indicherebbe l'esistenza di un *locus* genico principalmente coinvolto nell'insorgenza del MP, con chiari aspetti di ereditabilità mendeliana (molto probabilmente autosomico dominante a penetranza incompleta), corredato comunque da componenti multifattoriali (Cruz *et al.* 2008). Il prognatismo mandibolare non è limitato alle famiglie reali europee, oltre che nell'uomo, infatti, questa malformazione è stata riscontrata in specie animali appartenenti a diverse classi (*reptilia, aves, mammalia, amphibia*; Ritchie *et al.*, 1997) nonché alcune razze di cani brachicefali e nei conigli; in questi ultimi, l'analisi del *pedigree* avrebbe evidenziato per il MP una ereditarietà autosomica recessiva con penetranza incompleta (Huang *et al.* 1981).

Diversi tipi di deformità mandibolari, trattate più in dettaglio nel Paragrafo 1.3.1.1, sono state riportate anche per diverse specie di teleostei, sia di allevamento che selvatici.

Anche nel branzino (*D. labrax*) viene spesso riscontrato un fenotipo simile al prognatismo; in alcuni casi questa protrusione sembra essere dovuta ad ipertrofia della mandibola (Bargelloni L., comunicazione personale), in altre circostanze, invece, può essere spiegata come una compressione/iposviluppo della regione etmoidale e dell'arcata mandibolare superiore che rende l'arco mandibolare, normalmente

sviluppato, falsamente sporgente (condizione denominata “*pugheadness*” traducibile come “muso schiacciato”) (Mazurais *et al.*, 2009).

Nel branzino, questa malformazione è stata comunemente attribuita ad un eccesso di vitamina A nella dieta come anche l’assenza o l’eccesso di PUFA Ω 3 (*Ω -3 poly-unsaturated fatty acids*) (Mazurais *et al.*, 2009); sebbene fattori ambientali (per es. dieta, temperatura dell’acqua) sono considerati i principali agenti causali di deformità ossee nel branzino di allevamento, è stato recentemente dimostrato che anche fattori genetici giocano un ruolo di primo piano nello sviluppo di malformazioni alla spina dorsale (Bardon *et al.* 2009). Analogamente, analisi genetiche in una popolazione di *D. labrax* ha evidenziato un *bias* significativo nella frequenza del fenotipo prognato tra diverse famiglie *full-sib* (famiglie i cui membri condividono entrambi i genitori) sottoposte alle medesime condizioni di allevamento (Bargelloni L., comunicazione personale).

In questo studio, viene riportata l’analisi dei profili di espressione genica di esemplari normali ed esemplari affetti da MP. A questo scopo, è stata costruita una piattaforma *microarray* a oligonucleotidi, specifica per *D. labrax*, dapprima raccogliendo in un *database* tutti i trascritti unici di branzino, successivamente annotati, e disegnando poi due sonde (60meri) non sovrapposte per ciascun trascritto. Le sonde sono state sintetizzate *in situ* mediante la tecnologia Agilent *SurePrint*TM con la realizzazione, quindi, di una piattaforma *microarray* con più di 40.000 sonde (formato Agilent 4x44K).

Sebbene approcci simili siano già stati applicati per la costruzione di piattaforme *microarray* in altre specie di teleostei (vedi Paragrafo 3.1), questo studio è il primo a riportare la realizzazione di un *microarray* per lo studio dell’espressione genica in *D. labrax*; questo è piuttosto sorprendente visto l’importante ruolo che il branzino ricopre nel settore ittico europeo (vedi Paragrafo 1.2).

Nonostante l'allevamento di questa specie abbia visto un enorme incremento a partire dagli anni '80, diverse limitazioni ancora permangono, inclusa l'elevata frequenza di malformazioni craniche e scheletriche. Queste anomalie di sviluppo influiscono pesantemente sulla produzione del branzino in quanto gli esemplari affetti da deformità manifestano una *fitness* ridotta nonché una minor commerciabilità, richiedendo, quindi, un forte impiego di tempo e personale per la selezione manuale degli esemplari da rimuovere, con conseguenze economiche e di benessere animale rilevanti .

Come accennato in precedenza, evidenze preliminari indicano che il MP potrebbe avere una sostanziale componente genetica oltre alle già note cause ambientali; l'analisi dei profili di espressione di esemplari che manifestano questo fenotipo può essere considerato un approccio complementare alle analisi di *linkage* ad oggi in corso allo scopo di identificare quei *loci* coinvolti nella determinazione genetica di questo tratto.

Un approccio basato sull'analisi del trascrittoma, inoltre, può fornire una visione più ampia dei meccanismi molecolari che sottendono allo sviluppo di deformità craniche, cosa che potrebbe fare luce anche sui fattori ambientali che influenzano questa condizione.

In alcune specie, branzino incluso, sono già stati analizzati i profili di espressione genica, durante lo sviluppo/deformazione della mandibola in stadi ontogenetici precoci, di alcuni geni candidati (Villeneuve *et al.* 2005; Geay *et al.* 2009); lo scopo del presente studio è quello di analizzare l'espressione dell'mRNA in stadi più avanzati ed in regioni anatomiche circoscritte utilizzando una piattaforma *microarray* di neo-realizzazione. Allo stesso tempo, l'analisi dei profili di espressione caratterizzanti gli esemplari affetti da MP permette di valutare per la prima volta le *performance* della piattaforma stessa.

4.2 MATERIALI E METODI

4.2.1 Campioni biologici ed estrazione dell'RNA

Due diversi stadi di sviluppo di *D. labrax*; esemplari di 38 giorni (Stadio 38; lunghezza media 12 mm) e 58 giorni di età (Stadio 58; lunghezza media 16 mm), sono stati raccolti presso l' Impianto di Acquacoltura "Ca' Zuliani" (Pila di Porto Tolle) e sacrificati utilizzando un eccesso di anestetico.

Per quanto riguarda lo Stadio 38, la regione craniale (dalle labbra all'opercolo) è stata dissezionata rispettivamente da 15 esemplari normali e 15 esemplari affetti da prognatismo; le sezioni di cinque esemplari sono state poi impiegate per la creazione di tre *pool* indipendenti per ciascuna delle due condizioni.

Per quanto riguarda lo Stadio 58, invece, essendo gli esemplari di dimensioni maggiori, è stato possibile dissezionare solo l'arco mandibolare da un totale di 40 esemplari (20 normali e 20 affetti da prognatismo); analogamente allo Stadio 38, le sezioni di cinque esemplari sono state poi impiegate per la creazione di quattro *pool* indipendenti per ciascuna delle due condizioni.

Per ciascun *pool*, l'RNA totale è stato estratto utilizzando l' *RNAeasy Mini Kit* (Qiagen); la qualità dell'RNA è stata poi testata tramite elettroforesi su gel di agarosio all'1% contenente *SYBR Safe™ DNA Gel stain 10,000X* (Invitrogen™) e la concentrazione dell'estratto è stata determinata mediante lettura spettrofotometrica con lo strumento NanoDrop® ND-1000 (NanoDrop Technologies). La qualità e l'integrità dell'RNA è stata inoltre stimata con l'utilizzo del 2100 Bioanalyzer (Agilent Technologies); l'*RNA Integrity Number* (RIN), indice che fornisce una determinazione numerica e quindi oggettiva dell'integrità del campione, è stato, infatti, calcolato per ciascun estratto. Solo campioni dotati di RIN superiore a 8 sono stati utilizzati per i successivi esperimenti.

4.2.2 DLPD *database* e annotazione dei trascritti di *D. labrax*

Allo scopo di ottenere trascritti unici di branzino, sono state utilizzate tutte le sequenze derivanti da un pre-esistente *assembly* (versione 1.0 Adl1), scaricabile dal sito web del SIGENAE (<http://www.sigenae.org/>) all'inizio del presente studio (Febbraio 2008), costruito a partire da 14 librerie normalizzate di cDNA rappresentanti diversi tessuti (fegato, ovario, testicolo, osso/cartilagine, cervello/ghiandola pituitaria, cuore/vasi sanguigni, tessuto adiposo, rene cefalico, rene, branchie, intestino, milza, pelle, muscolo), una libreria di cDNA di stati larvali di branzino e due librerie sottratte derivanti da due infezioni sperimentali.

Questo *set* iniziale di 17.623 sequenze pre-assemblate, al quale sono state aggiunte 5.045 EST derivanti da una libreria di cDNA normalizzata di Corpuscoli di Stannius, 2.356 EST e 226 sequenze di mRNA depositate in GenBank, sono state assemblate utilizzando una strategia basata su analisi di Blast per l'identificazione delle sequenze (*e-value* soglia =1E-10) da includere nello stesso *cluster* e successivamente sull'utilizzo del programma Cap3 (Huang *et al.* 1999) per effettuarne l'assemblaggio e produrre le sequenze consenso.

Tutte le EST sono state sottomesse in NCBI, l'*accession number* in GenBank della sequenza che mostra la più elevata identità con ciascun *cluster* è riportata nel "*Platform data table*" (GPL9663) depositato in GEO.

Ricerche di similarità per ciascuno dei trascritti unici di branzino sono state effettuate tramite Blast; la procedura di annotazione ha previsto tre fasi:

1. Analisi di Blast, utilizzando l'opzione blastx (*e-value* soglia 1E-3), contro il *database* aminoacidico *non-redundant* (nr) di NCBI (*National Centre for Biotechnology Information*) nonché i *database* SwissProt e TrEMBL;

2. Analisi di Blast, utilizzando l'opzione blastn (*e-value* soglia 1E-5), contro il *database* nucleotidico *non-redundant* di NCBI;
3. Analisi di Blast, utilizzando l'opzione blastn (*e-value* soglia 1E-5), contro gli interi trascrittomi, depositati in ENSEMBL, di cinque specie di teleostei: zebrafish, stickleback, medaka, fugu e tetraodon.

Ove possibile, le associazioni di ciascun trascritti con i termini della *Gene Ontology* (GO ID), rispettivamente per “*Biological process*”, “*Molecular function*” e “*Cellular component*”, sono state ottenute tramite l'opzione blastx, implementata nel software Blast2GO (Götz *et al.* 2008), contro il *database* aminoacidico nr di NCBI.

Allo scopo di avere una visione di insieme del contenuto in GO, i risultati ottenuti dall'annotazione sono stati ricondotti all'interno dei termini contenuti nella “*Generic GO slim*” (<http://www.geneontology.org/GO.slims.shtml>) utilizzando il programma CateGORizer (Hu *et al.* 2008).

Tutti i trascritti unici di branzino e le annotazioni ad essi associate sono stati raccolti nel *database* DLPD (DLPD: *Dicentrarchus labrax* Padova), creato sulla base dell'ambiente BioMart. Questo *database* può essere interrogato utilizzando diversi filtri tra cui l'ID di ciascun *cluster*, la sua descrizione, il GO ID o una combinazione di questi criteri; noto l'ID del *cluster* è possibile, inoltre, risalire a diverse informazioni che comprendono la corrispondente sequenza consenso, la sua annotazione, e i presunti ortologhi umani nonché delle cinque specie di teleostei sopracitati (*G. aculeatus*, *O. latipes*, *D. rerio*, *T. nigroviridis*, and *T. rubripes*).

4.2.3 Disegno delle sonde ed esperimenti di *microarray*

Allo scopo di sviluppare una piattaforma *microarray* ad elevata densità, per ciascuno dei trascritti unici sono state disegnate due sonde non sovrapposte; il disegno è stato realizzato per mezzo dell'interfaccia Agilent eArray

(<https://earray.chem.agilent.com/earray/>) che ha sfruttato algoritmi di uso esclusivo per il designo di sonde costituite di 60 nucleotidi (60meri).

Le sonde sono state quindi sintetizzate *in situ*, utilizzando la tecnologia Agilent *SurePrint™*, per la costruzione di una piattaforma di formato 4x44K che prevede la sintesi di quattro *microarray*, ciascuno costituito di circa 44.000 sonde, in un'unica *slide*. Ciascun *array* include sonde di controllo positivo e negativo.

La marcatura e l'ibridazione di ciascun campione sono state eseguite in accordo al protocollo “*Agilent One-Color Microarray-Based Gene Expression Analysis*”; nel Paragrafo 3.2.3 è riportata una descrizione più dettagliata delle fasi di marcatura e ibridazione del campione.

4.2.4 Acquisizione ed analisi dei dati sperimentali

Lo scanner Agilent “*G2565BA DNA microarray scanner*” è stato utilizzato per la scansione degli *array* ad una risoluzione di 5 µm; ciascuna *slide* è stata sottoposta a una duplice scansione utilizzando due diversi livelli di sensibilità del laser (XDR Hi 100% and XDR Lo 10%).

Il *software Feature Extraction 9.5.1* è stato a questo punto utilizzato per l'analisi delle immagini ottenute; le due immagini generate per ciascuna *slide* sono state analizzate contemporaneamente ed i dati di espressione sono stati estrapolati e “puliti” dal *background* seguendo la procedura *standard* del programma. *Feature Extraction 9.5.1* fornisce anche tutta una serie di indicatori di qualità allo scopo di valutare la bontà e l'affidabilità dei valori di intensità stimati per ciascuno *spot*.

Tutte le sonde di controllo, ad eccezione degli *Spike-In*, sono state escluse dalle analisi successive; i valori di intensità degli *Spike-In* sono stati utilizzati per identificare la procedura di normalizzazione migliore per ciascuna serie di dati; dopo normalizzazione, infatti, l'intensità degli *Spike-In* dovrebbe essere uniforme tra i diversi esperimenti del

dataset. Le procedure di normalizzazione sono state eseguite utilizzando il programma statistico R (<http://www.r-project.org>); per le nostre serie di dati, l'approccio basato sulla normalizzazione quantile si è sempre dimostrato più performante della normalizzazione *cyclic lowess*; i dati normalizzati con l'approccio quantile sono stati quindi utilizzati per le analisi statistiche.

I test statistici implementati nel programma SAM (*Significance Analysis of Microarray*; Tusher *et al.* 2001) sono stati impiegati per identificare i geni differenzialmente espressi tra gli esemplari normali e prognati di ciascuno stadio. Per l'analisi dello Stadio 58, le sonde con una bassa proporzione di ibridazioni riuscite (più di due ibridazioni non riuscite tra le quattro repliche biologiche di ciascuna condizione) sono state eliminate dal *dataset*. La voce "*glsFound*" (corrispondente a 1 se lo *spot* ha un valore di intensità significativamente superiore al *background*, altrimenti equivalente a 0), restituita dal programma *Feature Extraction 9.5.1* in seguito all'estrapolazione dei dati di fluorescenza, è stata utilizzata per filtrare le sonde non attendibili (convenzionalmente *glsFound* = 0 è stato fatto corrispondere a "ibridazione non riuscita").

Allo scopo di valutare la precisione e la ripetibilità della piattaforma *microarray*, i coefficienti di correlazione di Pearson sono stati calcolati, sia tra esperimenti indipendenti che all'interno dello stesso, con il programma Statgraphics. Il test di correlazione non parametrica di Spearman, invece, è stato impiegato per valutare la correlazione tra i valori di espressione misurati rispettivamente con la piattaforma *microarray* o la metodica della *Real-time* RT-PCR; ciascun test è stato eseguito separatamente per ciascuna sonda utilizzando SPSS ver. 12.0.

4.2.5 Annotazione funzionale dei trascritti di *D. labrax*

Una annotazione funzionale dei geni differenzialmente espressi è stata eseguita utilizzando il server DAVID (*Database for Annotation, Visualization, and Integrated Discovery*; <http://david.abcc.ncifcrf.gov/>) (parametri di utilizzo gene count=4 and

ease=0.05). Poiché il *database* di DAVID contiene le annotazioni funzionali solo di un numero limitato di specie, è stato necessario associare ai trascritti di branzino (DLPD ID) un identificatore riconoscibile dal programma; nel nostro caso l' *Ensembl Human Gene ID* (ID dei geni umani depositati in Ensembl) e/o lo ZFIN-ID (ID dei geni di zebrafish depositati in ZFIN; http://zfin.org/cgi-bin/webdriver?MIval=aa-ZDB_home.apg).

Questa associazione è stata eseguita tramite diversi approcci:

1. Entrambe le opzioni di Blast, blastx e blastn, sono state utilizzate per la ricerca di *match* significativi tra le sequenze DLPD e rispettivamente le proteine o i trascritti umani depositati in Ensembl;
2. La stessa strategia di ricerca è stata diretta in primo luogo al proteoma e trascrittoma di *G. aculeatus* depositati in Ensembl e, successivamente, i *match* ottenuti tramite blastx (proteine di stickleback) o blastn (proteine predette dai trascritti di stickleback) sono stati a loro volta utilizzati in una seconda ricerca, tramite l'opzione blastp, delle corrispondenti proteine umane.
3. Una ricerca diretta, sempre utilizzando le opzioni blastn o blastx, contro tutte le proteine di zebrafish depositate in Ensembl
4. Una ricerca indiretta delle proteine di zebrafish dapprima passando attraverso i trascritti e le proteine di stickleback e successivamente identificando gli ortologhi di zebrafish, analogamente a quanto descritto al punto 2.

Uno schema riassuntivo dei quattro diversi approcci utilizzati è raffigurato in Appendice 4A.

Per concludere, gli *Ensembl Human Gene ID* (ENSG ID) e gli ZFIN-ID sono stati ottenuti direttamente dagli ID delle corrispondenti proteine depositate in Ensembl utilizzando BIOMART (<http://www.ensembl.org/biomart/martview/>).

4.2.6 Validazione dei dati di espressione mediante *Real-time* RT-PCR

Tredici geni *target* sono stati selezionati per l'analisi di espressione mediante *Real-time* RT-PCR; per ciascuno dei geni selezionati i *primer* gene-specifici sono stati disegnati col il *software* Primer 3 (<http://frodo.wi.mit.edu/primer3/>)

Per poter disegnare i *primer* a cavallo di un introne, le giunzioni esone-introne dei trascritti, ove possibile, sono state identificate in base all'allineamento con le sequenze genomiche corrispondenti, rese disponibili dal progetto di sequenziamento del genoma di branzino; altrimenti, sono state dedotte sulla base del confronto con le sequenze genomiche di geni omologhi in *T. nigroviridis*, *D. rerio* e *G. aculeatus*.

I campioni testati in *Real-time* RT-PCR sono gli stessi utilizzati negli esperimenti di *microarray*, incluso il calibratore. Un microgrammo di RNA totale di ciascun campione è stato retrotrascritto a cDNA utilizzando l'enzima Superscript II (Invitrogen™); una aliquota (2.5 µl) di cDNA diluito (2ng/µl) è stato quindi amplificato in un volume totale di 10 µl contenente 5 µl di *Platinum SYBR Green qPCR SuperMix-UDG 2X* (Invitrogen™) e 0,25 µl di ciascun *primer* gene-specifico (10 µM). Tutti gli esperimenti sono stati effettuati con lo strumento LightCycler® 480 (Roche Diagnostics); il protocollo di amplificazione ha previsto uno *step* iniziale di 2 min a 50°C e 2 min a 95°C seguito da 45 cicli di 10'' a 95°C e 30'' a 60°C.

Per valutare l'efficienza di ciascun saggio, curve standard sono state costruite amplificando diluizioni seriali (1:2) dello stesso cDNA utilizzato come calibratore (38d_N3). Il valore di Cp (*Crossing point*) è stato utilizzato per determinare la concentrazione relativa dei geni *target* in ciascun campione; tutte le quantificazioni sono state effettuate in duplicato e normalizzate al gene *reference*, MDH (DLPD06340), scelto poiché comunemente considerato gene *housekeeping* e perchè non mostrava alcuna variazione di espressione significativa né tra i due stadi di sviluppo né tra le due diverse condizioni (esemplari normali e prognati).

4.3 RISULTATI

4.3.1 Database DLPD

L'assemblaggio di tutte le sequenze disponibili di branzino ha prodotto un totale di 19.480 trascritti unici che sono stati raccolti nel *database* DLPD (*Dicentrarchus labrax* Padova); circa la metà di questi (9.497) ha mostrato un *match* significativo, in seguito a blastx, contro una proteina nota depositata nei *database* pubblici SwissProt, TrEMBL o NCBI (*non redundant* aminoacidico). Un diagramma di Venn rappresentante il numero di match ottenuti con ciascun *database* è rappresentato in Figura 4.1. Un ulteriore incremento nel numero di sequenze DLPD annotate è stato ottenuto mediante blastn contro il *database non redundant* (nr) nucleotidico di NCBI; questo approccio ha prodotto un *match* significativo per 2.511 ulteriori sequenze che non mostravano corrispondenza con alcuna delle proteine depositate nei *database* proteici sopracitati (Figura 4.1). La somma dei due approcci ha portato il numero totale di trascritti di branzino associati ad una proteina o un trascritto noti a 12.008 (63%).

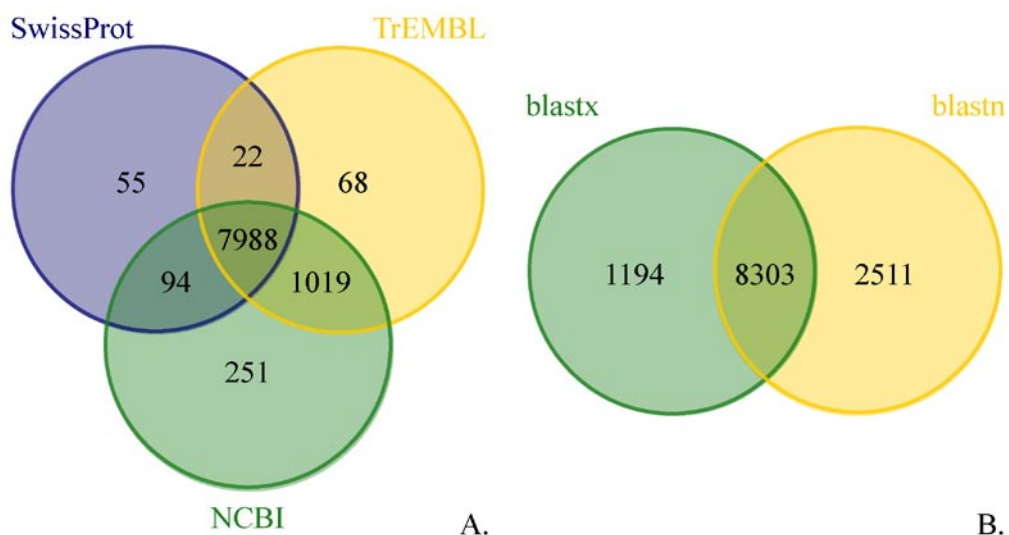


Figura 4.1: A. Diagramma di Venn raffigurante il contributo di ciascun *database* proteico all'annotazione dei trascritti unici di branzino mediante l'approccio basato su analisi di blastx. B. Diagramma di Venn raffigurante il contributo dei due diversi approcci, blastx contro i *database* proteici e blastn contro il nr nucleotidico di NCBI, all'annotazione dei trascritti di branzino.

Ricerche teleosteo-specifiche contro i trascrittomi delle cinque specie depositate in Ensembl (*G. aculeatus*, *O. latipes*, *D. rerio*, *T. nigroviridis*, e *T. rubripes*) hanno permesso di identificare un totale di 9.506 sequenze DLPD che mostravano una corrispondenza significativa con almeno una delle specie (Tabella 4.1); 4.504 di queste mostravano un *match* significativo con tutti i cinque *database*.

Tabella 4.1 Numero di trascritti DLPD che mostrano *match* significativi in ciascuno dei cinque trascrittomi di teleostei depositati in Ensembl.

Trascrittomi in Ensembl	Sequenze DLPD con <i>match</i> significativo (% sequenze DLPD)
<i>Gasterosteus aculeatus</i> (stickleback)	8448 (44.4%)
<i>Oryzias latipes</i> (medaka)	7004 (36.8%)
<i>Takifugu rubripes</i> (fugu)	7774 (40.8%)
<i>Tetraodon nigroviridis</i> (tetraodon)	7739 (40.6%)
<i>Danio rerio</i> (zebrafish)	5579 (29.3%)

Un'ulteriore analisi è stata eseguita utilizzando il *software* Blast2GO, il quale ha permesso di associare uno o più termini GO a 4.692 trascritti DLPD; a 3.267 di questi è stato associato almeno un termine GO per “*Biological process*” (BP), a 2.858 almeno un termine per “*Cellular Component*” (CC) e a 3.879 almeno un termine per “*Molecular Function*” (MF). Il numero di termini di *Gene Ontology* unici rappresentati nel database DLPD sono risultati essere rispettivamente 1.541 per BP, 381 per CC e 998 per MF; una versione semplificata di questi GO, basata sulle voci raccolte nella “*Generic GO Slim*”, mostra un totale di 47 termini parentali per BP, 28 per CC e 36 per MF (vedi Appendice 4C).

Tutti i trascritti unici di *Dicentrarchus labrax* e le annotazioni ad essi associate, i GO ID e i potenziali geni omologhi di altre specie di pesci modello, sono state depositate in un *database* dedicato e di pubblico accesso: *DLPD database* (<http://enne.cribi.unipd.it:5556/biomart/martview>).

4.3.2 Valutazione della qualità della piattaforma microarray

Il disegno delle sonde a oligonucleotidi (due sonde non sovrapposte per ciascun trascritto) è stato completato con successo per 19.035 trascritti DLDP; le sequenze di ciascuna sonda e altri dettagli sulla piattaforma *microarray* possono essere visionati nel “*Platform data table*” depositato in GEO (*accession number* GPL9663).

La qualità di ciascuna sonda è stata testata sperimentalmente in base al numero di ibridazioni riuscite su un totale di 14 esperimenti (sei ibridazioni indipendenti per lo Stadio 38 e otto per lo Stadio 58); le ibridazioni sono state considerate “riuscite” solo per le sonde alle quali era associato un *glsFound* = 1 dal programma *Feature Extraction 9.5.1*. Solo tre sonde (0,008%) non hanno mai mostrato un’intensità di segnale superiore al rumore di fondo mentre per 35.957 oligonucleotidi, corrispondenti al 94.4% del numero totale di sonde, le ibridazioni sono riuscite in almeno la metà degli esperimenti (Figura 4.2).

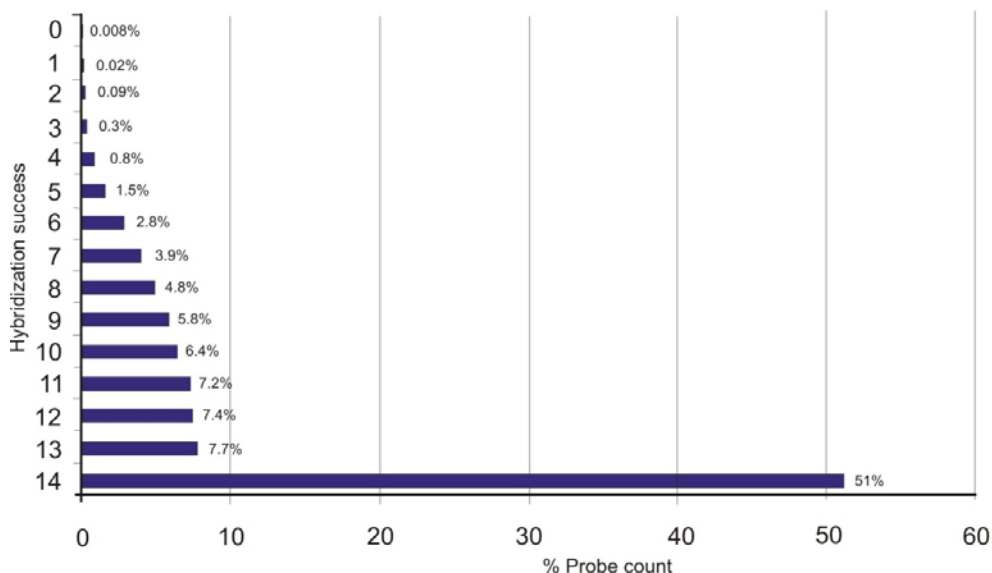


Figura 4.2: Analisi della distribuzione del successo di ibridazione di ciascuna sonda su un totale di 14 esperimenti. Per ciascuna sonda è stato calcolato il numero di ibridazioni riuscite mediante il programma *Feature Extraction 9.5.1*. Asse Y: numero di ibridazioni riuscite in 14 esperimenti (0 corrisponde al numero di sonde che non hanno mai ibridato, 14 corrisponde al numero di sonde per le quali le ibridazioni sono sempre riuscite. Asse X: numero di sonde che ricade in ciascun gruppo (0-14).

Uno dei requisiti più importanti in un esperimento di *microarray* è la riproducibilità; allo scopo di valutare la riproducibilità e la precisione del protocollo sperimentale e della piattaforma stessa, i sei *pool* dello Stadio 38 sono stati considerati repliche biologiche.

Il grado di concordanza tra le repliche è stato quindi calcolato tramite la stima del coefficiente di correlazione di Pearson per l'intero *set* di valori di fluorescenza. Per ciascun confronto a coppie dei sei esperimenti, i coefficienti di correlazione sono risultati essere sempre significativi (p -value <0,01) e mai inferiori allo 0,99 (Tabella 4.2) dimostrando, quindi, una correlazione quasi perfetta tra le repliche ed una buonissima ripetibilità della piattaforma stessa.

Tabella 4.2 Coefficienti di correlazione di Pearson calcolati, per ciascun confronto a coppie, nelle sei repliche biologiche dello Stadio 38

	38d_N3	38d_N4	38d_N6	38d_P3	38d_P5	38d_P6
38d_N3	1.00					
38d_N4	0.996**	1.00				
38d_N6	0.996**	0.998**	1.00			
38d_P3	0.991**	0.994**	0.991**	1.00		
38d_P5	0.993**	0.996**	0.996**	0.994**	1.00	
38d_P6	0.994**	0.996**	0.997**	0.993**	0.996**	1.00

** p -value <0,01

La piattaforma *microarray* di orata è caratterizzata dalla presenza di due sonde per ciascun trascritto; queste due sonde si appaiano alla sequenza genica in due regioni non sovrapposte; in particolare, la prima sonda (Sonda_1) è progettata per essere il più vicino possibile al 3' di ciascun trascritto. Alla luce di questo, la variabilità tra le due sonde disegnate per ciascun trascritto è stata valutata utilizzando il rapporto tra i due valori di intensità (*fold-change*, FC) come misura della differenza di segnale. Questo rapporto dovrebbe dare una differenza il più vicina possibile (se non uguale) a uno (1); a meno che non intervengano fattori tecnici (per es. scarsa qualità delle sonde) o biologici

(per es. forme di *splicing* alternativo) a determinare una deviazione rispetto al valore atteso.

In Figura 4.3, ciascun *boxplot* mostra la distribuzione dei *fold-change* osservati tra Sonda_1 e Sonda_2 in ciascuno degli esperimenti eseguiti sui *pool* di teste dello Stadio 38; come atteso, in tutti gli esperimenti i rapporti tra le intensità delle due sonde per lo stesso gene mostrano una distribuzione simmetrica e centrata sull'uno. Le coppie di sonde mostrano, infatti, piccole differenze in termini di intensità di segnale, con il 68% dei trascritti caratterizzati da un FC inferiore a 2.

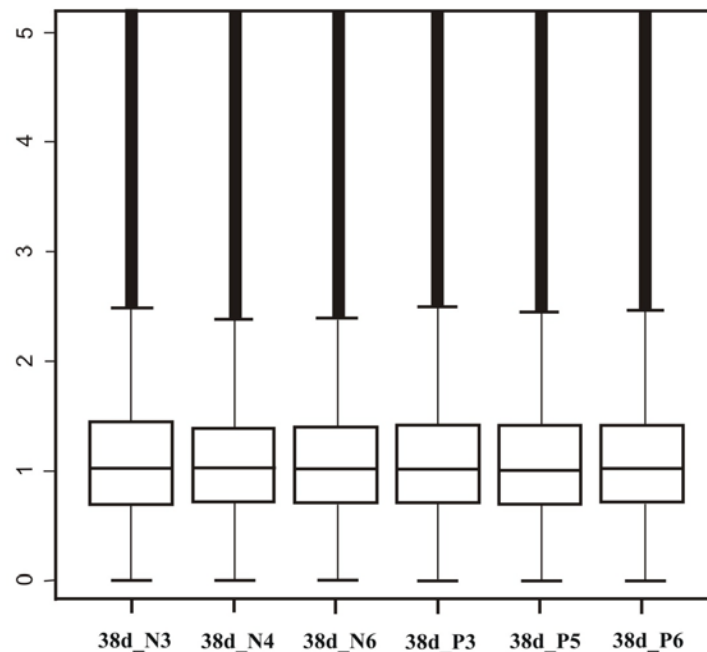


Figura 4.3: Boxplot della distribuzione dei *fold-change* osservati (asse y), tra le due sonde (Sonda_1 e Sonda_2) di ciascun trascritto, in tutte le repliche biologiche dello Stadio 38.

Per confermare ulteriormente la complessiva concordanza tra le coppie di sonde, è stato calcolato il coefficiente di correlazione di Pearson, tra la Sonda_1 e la Sonda_2 di ciascun trascritto, per tutti i sei esperimenti dello Stadio 38; in tutti i confronti a coppie considerati, la correlazione è risultata essere sempre significativa ($p < 0,01$) con coefficienti mai inferiori a 0,93.

4.3.3 Analisi dei dati di espressione genica

I dati di fluorescenza grezzi e quelli normalizzati, sia degli esperimenti *microarray* sullo Stadio 38 che sullo Stadio 58, sono stati depositati nel database GEO con *accession number* rispettivamente GSE19041 e GSE19001.

Analisi statistiche, utilizzando il programma SAM, sui dati normalizzati sono state effettuate, allo scopo di identificare i geni differenzialmente espressi tra esemplari normali e prognati, separatamente per lo Stadio 38 (intere regioni craniche) e lo Stadio 58 (arcate mandibolari). L'analisi statistica effettuata sullo Stadio 38 non ha permesso di identificare alcun gene significativo; al contrario, anche dopo la completa esclusione delle sonde con scarso successo di ibridazione (6.404 sonde), l'analisi dei dati per lo Stadio 58 ha rivelato un numero consistente di geni differenzialmente espressi tra mandibole normali e deformi. Dopo aver filtrato le sonde non conformi, i dati normalizzati di 17.696 (93%) trascritti DLPD, rappresentati da una (3.722) o due sonde (13.972), sono stati sottoposti ad analisi statistica tramite "SAM due classi"; utilizzando come soglia un FDR del 5%, è stata ottenuta una lista di 333 sonde, corrispondenti a 242 trascritti di *D. labrax* (per 91 trascritti sia la Sonda_1 che la Sonda_2 sono risultate essere differenzialmente espresse mentre i rimanenti 151 trascritti erano rappresentati da solo una delle due sonde).

Tutti i 242 trascritti di branzino sono risultati essere sottoespressi nei *pool* di mandibole affette da prognatismo rispetto ai controlli, con FC che variano da -1,12 a -15,7 volte; 70 di questi mostrano una sottoespressione superiore alle 2 volte (per due trascritti superiore alle 10 volte).

A 122 dei 242 trascritti è stato possibile associare una annotazione e, tra i 70 geni che mostrano i livelli più elevati di *fold-change*, la metà ha un *match* significativo con una proteina/sequenza nucleotidica nota. Una porzione rilevante delle proteine omologhe a questi trascritti DLPD sembrano essere coinvolte nella funzione sinaptica

(*Synaptophysin*, *SNAP-25a*, *N-ethylmaleimide sensitive fusion protein/vesicle-fusing ATPase*, *Syntaxin binding protein 1*, *Neuroplastin*), nello sviluppo neuronale (*Microtubule-associated protein 2*, *Contactin-1*, *Collapsin response mediator protein-1*, *Neurogenic differentiation factor 2*, *Chondroitin sulfate proteoglycan 5 Precursor*) o, più in generale, risultano essere specifiche del sistema nervoso (*Fatty acid-binding protein brain isoform*, *Visinin-like protein 1*, *Beta-synuclein*, *Plasticin/peripherin*, *Secretogranin II*).

Questa osservazione viene ulteriormente confermata andando ad analizzare l'intero *set* dei geni differenzialmente espressi, considerando, quindi, anche i trascritti con *fold-change* < 2 (per es. *Neuronal membrane glycoprotein M6-a*, *Carboxypeptidase E*, *Neural Wiskott-Aldrich syndrome protein*, *Schwannomin-interacting protein 1*, *Synapsin-2B*, *Neuroserpin*, *Ephrin type-B receptor 3*, *Metabotropic glutamate receptor 8*, *RUFY3*, *GMP-PDE delta*); un ulteriore gene che mostra una sostanziale sottoespressione (- 4,6 volte), *Zinc finger protein of the cerebellum 3 (Zic3)*, è stato riportato giocare un ruolo critico nello sviluppo del tessuto nervoso, nell'organizzazione dell'asse destra-sinistra nonché nello sviluppo dei somiti.

4.3.4 Annotazione funzionale dei geni differenzialmente espressi

Allo scopo di ottenere una più sistematica interpretazione funzionale del *set* di geni risultati essere differenzialmente espressi tra gli esemplari prognati e i controlli, è stata eseguita un'analisi di arricchimento utilizzando il *software* bioinformatico DAVID.

Seguendo i diversi approcci riportati nel Paragrafo 4.2.5, le sequenze DLPD sono state associate ad "identificatori", *Ensembl Gene ID* umani e/o *ZFIN IDs*, con un numero di corrispondenze, tra le sequenze di branzino e le corrispondenti di uomo/zebrafish, variabile a seconda del metodo impiegato (Appendice 4A).

Gli “identificatori” corrispondenti ai trascritti di branzino differenzialmente espressi sono stati utilizzati come “*gene list*” mentre, quelli corrispondenti all’intero *set* di sequenze DLPD, sono stati utilizzati come “*background*” in DAVID. Il *software* permette una annotazione funzionale dei geni significativi attraverso l’interrogazione di un *database* biologico integrato che contiene più di 40 categorie di annotazione. In questo studio vengono riportati i risultati ottenuti dal confronto diretto delle sequenze DLPD con il trascrittoma umano (vedi Appendice 4A).

Un totale di 60 omologhi umani dei trascritti DLPD differenzialmente espressi erano presenti nel *database* di DAVID, secondo l’analisi di arricchimento 14 termini GO per “*Biological process*” (GO_BP) sono risultati essere significativamente sovrarappresentati (Tabella 4.3).

Tabella 4.3: Termini GO significativamente rappresentati all’interno del *set* di geni risultati essere differenzialmente espressi tra mandibole di esemplari prognati e normali.

Categoria	Termini GO	Tot	p-value	FDR
GOTERM BP_ALL	GO:0050877~neurological system process	8	0.006574	11.51896
	GO:0048856~anatomical structure development	14	0.026081	38.75987
	GO:0003008~system process	9	0.022777	34.78694
	GO:0032501~multicellular organismal process	23	2.19E-04	0.406379
	GO:0050789~regulation of biological process	24	0.01166	19.5573
	GO:0007154~cell communication	19	0.034946	48.31694
	GO:0007399~nervous system development	7	0.039474	52.63508
	GO:0007275~multicellular organismal development	15	0.011784	19.74412
	GO:0007268~synaptic transmission	6	0.002618	4.747339
	GO:0006811~ion transport	7	0.022343	34.24804
	GO:0048731~system development	13	0.013046	21.62432
	GO:0001505~regulation of neurotransmitter levels	4	0.002972	5.3725
	GO:0019226~transmission of nerve impulse	6	0.005133	9.107265
	GO:0065007~biological regulation	26	0.011622	19.49988
GOTERM CC_ALL	GO:0005856~cytoskeleton	10	0.013466	18.45919
GOTERM MF_ALL	GO:0008092~cytoskeletal protein binding	6	0.046655	56.07969
UP_TISSUE	Brain	39	1.78E-04	0.254897

Circa la metà dei geni (26, il 43%) rientranti in questi GO_BP giocano un ruolo nella regolazione dei processi biologici (GO:0050789) e 14 sono coinvolti nello sviluppo delle strutture anatomiche (GO:0048856); entrando più nello specifico, cinque termini sono direttamente correlati allo sviluppo e alla funzione del sistema nervoso (*neurological system process, synaptic transmission, regulation of neurotransmitter levels, nervous system development, transmission of nerve impulse*).

Due termini GO per “*Cellular component*” (GO:0005856 “*cytoskeleton*”, P=0.01; GO:0019717 “*synaptosome*”, P=0.02) sono stati evidenziati da DAVID come sovrarappresentati; entrambi sono correlabili all’unica voce evidenziata per “*Molecular function*” (GO:0008092 “*cytoskeletal protein binding*”) in quanto le proteine citoscheletriche risultano coinvolte nella genesi e nella funzione delle sinapsi.

Un’ulteriore evidenza del coinvolgimento sia dei processi che delle componenti del sistema nervoso è data dalla tessuto-specificità dei geni differenzialmente espressi; 39 geni infatti mostrano la loro espressione a livello del cervello.

4.3.5 Validazione mediante *Real-time* RT-PCR

Allo scopo di validare le *performance* della piattaforma *microarray*, saggi di espressione in *Real-time* RT-PCR sono stati effettuati su un *set* di 13 geni, risultati essere differenzialmente espressi nello Stadio 58, scelti tra i geni annotati e che rappresentassero l’intera gamma di *fold-change* riscontrati (1,06-13 volte).

Allo scopo di aumentare il numero di dati su cui effettuare le analisi di correlazione tra *microarray* e qPCR, i campioni appartenenti allo Stadio 38 sono stati inclusi nelle analisi e i *fold-change* per i 13 geni target selezionati, sono stati quindi calcolati tra Stadio 38 e Stadio 58 (una normalizzazione quantile è stata effettuata *ad hoc* su tutti i dati di *microarray*, 14 esperimenti).

Complessivamente, una correlazione statisticamente significativa tra qPCR e *microarray* è stata ottenuta confrontando i valori di espressione di ciascun gene *target* in tutti i campioni (Tabella 4.4). Nove geni hanno mostrato elevati coefficienti di correlazione (Spearman rho > 0,8) per entrambe le sonde (p < 0,01) e altri 3 trascritti hanno riportato una significativa correlazione positiva (0,7 < rho < 0,8 con p < 0,01). Non è stata, invece, riscontrata correlazione tra le due metodiche per il gene *Cyclin Dependent Kinase* (CDK), caratterizzato da un *fold-change* inferiore alle 1,2 volte.

Tabella 4.4: Correlazione tra i valori di espressione rilevati dalla *Real-time* RT-PCR e le due sonde *microarray* per ciascun trascritto *target*.

Gene	Rho di Spearman		
	Sonda_1/Sonda_2	qPCR/Sonda_1	qPCR/Sonda_2
Chondroitin sulfate proteoglycan 5 Precursor	0.867**	0.878**	0.912**
Calcitonin	0.887**	0.796**	0.846**
Stathmin	0.978**	0.930**	0.974**
Tetraspanin	0.991**	0.758**	0.734**
Peripherin	0.979**	0.951**	0.895**
Brain specific fatty acid protein	0.978**	0.960**	0.974**
Neurogenic differentiation factor	0.863**	0.830**	0.907**
Stromal cell derived receptor 1β	0.958**	0.762**	0.846**
Cyclin dependent kinase	0.952**	0.162	0.086
Synaptosomal associated protein	0.986**	0.902**	0.874**
Pleiotropin	0.979**	0.958**	0.944**
GRAM-domain containing protein 2 (splicing variant 1)	0.900**	0.934**	0.855**
GRAM-domain containing protein 2 (splicing variant 2)	0.991**	0.855**	0.890**

** p < 0,01

Per osservare la concordanza di direzione nel cambiamento di espressione tra le due metodiche, sono stati determinati i *fold-change* rilevati dai saggi in *Real-time* RT-PCR e da entrambe le sonde *microarray*. I *fold-change* sono stati calcolati come il rapporto delle intensità di segnale medie tra prognati e controlli rispettivamente nello Stadio 38 e nello Stadio 58 come anche il rapporto delle intensità di segnale medie tra i due stadi di sviluppo.

Per tutti i geni *target*, la direzione del cambiamento di espressione è risultata concorde tra qPCR e i dati di *microarray* (Figura 4.4). Per entrambe le sonde, i risultati ottenuti

dalla regressione lineare tra i *fold-change* stimati rispettivamente dalla qPCR e dalla piattaforma *microarray* dimostrano una forte correlazione positiva (Spearman rho > 0.92, *p-value* <0.01) tra le due tecnologie.

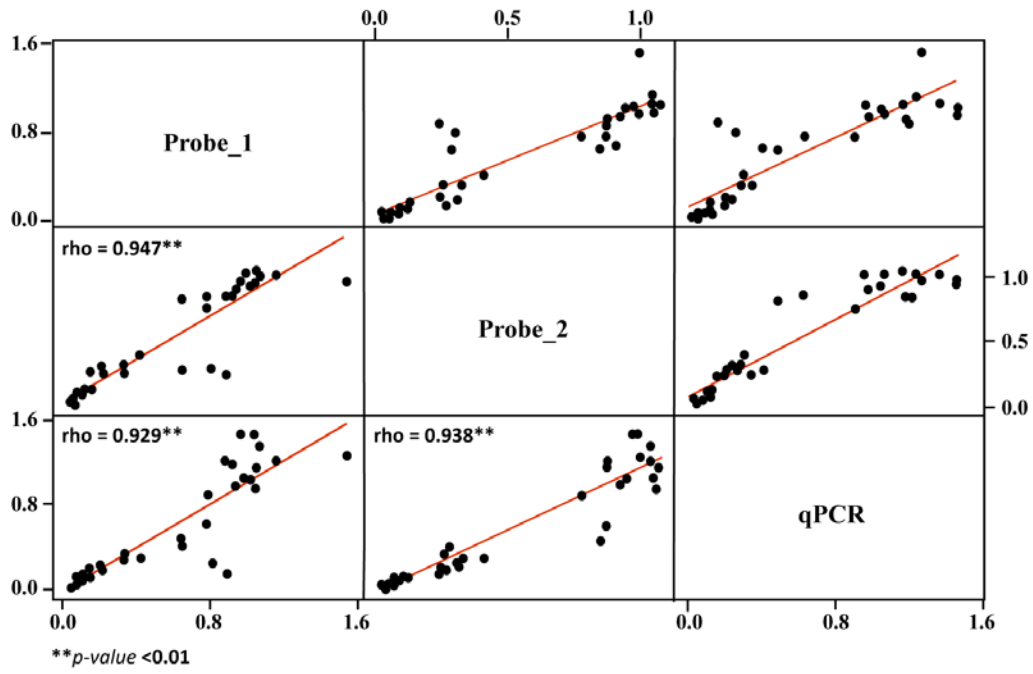


Figura 4.4: Correlazione tra i *fold-change* stimati da Sonda_1 (Probe_1), Sonda_2 (Probe_2) e qPCR

4.4 DISCUSSIONE

Sulla base dei risultati ottenuti in questo studio emergono quattro punti principali; il primo riguarda l'elevato livello di flessibilità, affidabilità e riproducibilità della piattaforma *microarray* sviluppata per *D. labrax*. Il disegno degli oligonucleotidi si è dimostrato essere fattibile per quasi tutti i trascritti disponibili per questa specie e quasi tutte le sonde hanno mostrato un buon successo di ibridazione fornendo la più completa rappresentazione del trascrittoma, ad oggi noto, di branzino. L'analisi delle repliche biologiche ha poi evidenziato coefficienti di correlazione estremamente elevati, dimostrando un'ottima riproducibilità dei dati di *microarray*. Come riportato da studi precedenti (Ferraresso *et al.* 2008; Millán *et al.* 2009), l'uso dell'approccio a un colore non ha avuto apparentemente alcun effetto negativo sulla qualità dei dati, al contrario, permette di semplificare il disegno sperimentale nonché l'analisi di nuovi campioni. Per concludere, la validazione mediante *Real-time* RT-PCR, metodica di quantificazione dell'espressione genica che si basa su un diverso approccio tecnico, ha confermato i risultati di *microarray* con l'eccezione di un unico gene, *Cyclin dependent kinase*, che ha mostrato una correlazione positiva ma non statisticamente significativa tra le due metodiche (Tabella 4.4). Questo è probabilmente dovuto alla differenza di espressione estremamente bassa di questo gene (*fold-change* medio stimato dal *microarray*: 1,3). La scarsa correlazione tra *microarray* e *Real-time* RT-PCR per i geni che esibiscono piccole differenze di espressione (< 1,4 volte) è, infatti, comunemente riportato, tant'è che un *fold-change* di 2 è considerato la soglia al di sotto della quale le due tecnologie perdono corrispondenza (Morey *et al.* 2006).

Il secondo punto riguarda l'annotazione del trascrittoma di *D. labrax*; dal momento che lo sviluppo e la validazione di piattaforme *microarray* personalizzate è diventata una procedura semplice, un'adeguata classificazione funzionale dei trascritti rappresentati

nel *microarray* è di sicuro più difficile da ottenere, in particolar modo per teleostei non-modello.

Le tecnologie *high-throughput* hanno determinato lo sviluppo di diversi approcci piuttosto esaurienti per l'analisi funzionale di liste di geni/proteine differenzialmente espressi; uno dei principali risulta essere la *Gene Ontology* (Ashburner *et al.* 2000). Questi metodi offrono dei vantaggi importanti, nonostante l'annotazione basata sulla *gene ontology* abbia dei limiti risaputi (Rhee *et al.* 2008). Quando questi approcci vengono estesi alle specie non-modello, la classificazione dei diversi geni è basata ancor più sull'ipotetica omologia con organismi meglio caratterizzati piuttosto che su evidenze sperimentali dirette. A questo proposito ci sono due principali fonti di errore: la prima si riferisce al problema di identificare reali relazioni ortologhe (Kuzniar *et al.* 2008), la seconda è dovuta al fatto che l'omologia di sequenza non sempre implica un'omologia funzionale, questo a causa delle differenze intrinseche tra le due specie comparate. Nel caso dei teleostei, l'estrapolazione di annotazioni funzionali da specie modello appartenenti alla classe dei mammiferi pone un problema ulteriore causato dalla duplicazione dell'intero genoma (WGD; *Whole Genome Duplication*) alla quale è stato sottoposto l'antenato dei pesci teleostei dopo la separazione dalla linea filetica dei tetrapodi (Kasahara *et al.* 2007). La WGD ha determinato la presenza, nei teleostei, di due copie per ciascuno dei geni presenti nei vertebrati superiori; una parte delle quali è stata ritenuta per mezzo di diversi meccanismi (per es. neo-funzionalizzazione, sub-funzionalizzazione, inerzia), mentre le altre copie sono state perse per mezzo di delezioni o pseudogenizzazioni (Kassahn *et al.* 2009).

Malgrado queste limitazioni, l'annotazione funzionale per mezzo dell'omologia di sequenza rimane il metodo migliore a disposizione per le specie non-modello; nel caso di *D. labrax*, una buona percentuale (>60%) dei trascritti ha mostrato un *match* significativo con un gene/proteina noti, percentuale nel complesso superiore a quelle

raggiunte in studi simili condotti su altre specie di teleostei, come per esempio il 40% ottenuto per l'orata (Ferraresso *et al.* 2008), il 50,7% del rombo (Pardo *et al.* 2008), il 40,6% della sogliola (Cerdà *et al.* 2008) e il 51% del pesce gatto (Liu *et al.* 2008), e comparabile a quanto riportato per l'halibut (60%; Douglas *et al.* 2007). A una proporzione decisamente più piccola di trascritti DLPD (circa 4.500) è stato possibile associare un termine GO, probabilmente come conseguenza dei criteri più stringenti utilizzati dal programma Blast2GO e della mancanza di annotazione GO per una parte delle sequenze proteiche corrispondenti ai trascritti DLPD.

Nonostante questo, sono state comunque raccolte un numero sufficiente di informazioni per costruire un *background* personalizzato, che tiene quindi conto del fatto che solo una parte del trascrittoma viene rappresentato nella serie di dati, e per identificare i processi biologici che risultano particolarmente rappresentati all'interno della lista dei geni differenzialmente espressi.

Un secondo approccio di annotazione per mezzo di similarità di sequenza, diretta o indiretta, con un trascrittoma modello (vedi Paragrafo 4.3.4) sembra portare a informazioni comparabili, con più di 6.000 potenziali omologhi umani e 5.000 di zebrafish; di questi, rispettivamente 5.654 e 4.150 erano corrispondenti a una voce riconosciuta da DAVID. L'utilizzo del trascrittoma umano come riferimento ha fornito un numero maggiore di annotazioni funzionali significative rispetto a zebrafish, anche se è da tenere a mente che l'utilizzo di annotazioni basate sul genoma umano può essere soggetto ai rischi sovradescritti.

Se la maggior parte delle sequenze DLPD sono state associate a un gene noto, esiste una sostanziale minoranza di trascritti "silenti", per i quali cioè non è stato possibile trovare alcun *match* significativo con nessuno dei *database* di sequenze esplorati; tuttavia, anche se non associati ad alcun gene noto, questi trascritti possono comunque fornire delle informazioni utili. Per esempio, una buona proporzione (190 su 242 con un *e-*

value soglia di 1E-5, o 169 su 242 con un *e-value* soglia di 1E-10) dei geni differenzialmente espressi nelle mandibole deformi, trova un *match* significativo su una specifica regione del genoma di stickleback; l'elevato livello di conservazione di ampi frammenti del genoma tra *D. labrax* e *G. aculeatus* (Reinhardt R., comunicazione personale) e la disponibilità di una mappa fisica del genoma di branzino (Galibert F., comunicazione personale) potrà in futuro permettere l'uso dei trascritti omologhi ai DLPD come candidati "posizionali" per identificare i *loci* coinvolti nella determinazione genetica del prognatismo mandibolare.

Il 40% dei trascritti "silenti" probabilmente è costituito da diversi tipi di trascritti, come per esempio 5'- e 3'-UTR (*UnTranslate Region*), o isoforme alternative di *splicing* che non sono state ancora caratterizzate in altre specie; una frazione di questi trascritti, poi, potrebbe appartenere alla categoria in continua espansione degli RNA non codificanti, che sembra ricoprire una parte sempre crescente del genoma animale (Carninci, 2007; Jacquier, 2009).

La quarta e ultima evidenza raccolta nel presente studio è stata l'identificazione di un *set* di trascritti differenzialmente espressi nelle mandibole di branzini affetti da prognatismo rispetto ai controlli. I difetti dello sviluppo sono generalmente considerati agire precocemente nell'ontogenesi; per esempio, mutazioni a carico di geni regolatori essenziali iniziano ad esercitare il loro effetto nelle fasi precoci dello sviluppo cranio-facciale (per es. *Tbx22* nei mammiferi e in zebrafish; Jezewski *et al.* 2009 ed *Endothelin1* in alcuni vertebrati; Miller *et al.* 2003) mentre il trattamento precoce di larve (6-9 giorni dopo la schiusa) di passera giapponese (***Paralichthys olivaceus***) con agonisti del recettore per l'acido retinoico determina l'insorgenza di deformazioni alla mascella inferiore (Haga *et al.* 2002).

In questo studio, alterazioni nei profili di espressione sono stati osservati in stadi più tardivi, queste potrebbero essere effetti "a valle" di cambiamenti di espressione

precedenti a carico di geni gerarchicamente superiori, oppure, perturbazioni nell'espressione genica che iniziano in uno stadio meno differenziato e che vengono mantenute nel tempo. E' da sottolineare però che, nel confronto tra individui normali e prognati, non sono state osservate differenze statisticamente significative dei profili di espressione in esemplari di 38 giorni; per questo stadio, tuttavia, sono state analizzate intere regioni craniche, fattore che potrebbe aver determinato una minor sensibilità nella rilevazione di potenziali alterazioni nell'espressione genica.

L'ampio *set* di trascritti sottoespressi nell'arcata mandibolare di esemplari prognati suggerisce che il MP osservato in questo studio potrebbe essere distinto dalla condizione nota col nome di "*pugheadness*" che è considerata essere il risultato di un iposviluppo della mascella superiore con una conseguente protrusione, solo apparente, della mandibola, normalmente sviluppata.

Come in molte altre specie di teleostei, lo sviluppo mandibolare in *D. labrax* è caratterizzato dalla presenza di componenti cartilaginee; la cartilagine di Meckel, che si sviluppa precocemente e successivamente regredisce ed ossifica, e l'osso dentale che invece appare più tardi (25 giorni dopo la schiusa) e va incontro ad ossificazione diretta (Kuzir *et al.* 2004); entrambe queste componenti ricevono un importante contributo di cellule originarie della cresta neurale (Calloni *et al.* 2009). Uno dei ruoli principali nel processo di ossificazione della cartilagine di Meckel è ricoperto dalla via di segnale dei geni *Hedgehog*, e appare essere conservato tra pesci e mammiferi, mentre i meccanismi molecolari che controllano lo sviluppo dell'osso dentale sono meno caratterizzati e sembrano essere, almeno in parte, distinti da quelli osservati per la cartilagine di Meckel (Abzhanov *et al.* 2007). Come si correlano questi meccanismi ai geni differenzialmente espressi identificati in questo studio? L'annotazione funzionale di questi trascritti ha evidenziato un significativo arricchimento in geni, di tipo regolatorio, coinvolti principalmente nello sviluppo di strutture anatomiche; una valutazione "gene per gene"

ha confermato che alcuni di questi possono giocare un ruolo nella formazione del tessuto osseo. Per esempio, il gene SOX4, espresso ad elevati livelli nei condriti ipertrofici durante l'ossificazione endocondrale (Reppe *et al.* 2000), è stato riportato agire in relazione all'ormone paratiroideo (PTH) e a proteine PTH-correlate (PTH-rp) in cellule osteoblasto-simili; la pleiotropina (o HB-GAM) è ampiamente espressa nell'osso, dove sembra giocare un ruolo nello sviluppo e rimodellamento del tessuto osseo stesso (Imai *et al.* 2009), ed infine, il recettore per l'acido retinoico RXR γ forma eterodimeri con i recettori per la vitamina D, i PPARs (*Peroxisome proliferator-activated receptors*) e i RARs (*Retinoic acid receptors*); tutti questi recettori nucleari sono in qualche modo coinvolti nella formazione dell'osso e, più in generale, nel controllo della crescita ossea.

E' da sottolineare anche che la vitamina A e i suoi metaboliti, agendo attraverso i RARs, sono stati riportati causare deformità cranio-facciali in pesci di allevamento (Mazurais *et al.* 2009); in particolare, alti livelli di vitamina A nella dieta aumenta significativamente la frequenza di deformità craniche (iposviluppo dell'arco mandibolare) in branzino (Mazurais *et al.* 2009) come anche in altre specie di teleostei (Haga *et al.* 2002; Takeuchi *et al.* 1998).

Lo sviluppo del tessuto osseo non è, tuttavia, il processo biologico più rappresentato all'interno del *set* di geni differenzialmente espressi negli esemplari affetti da prognatismo mandibolare; una porzione sostanziale di questi, infatti, sembrano influire sulla crescita e funzione nervosa. Questa evidenza può essere spiegata da un cambiamento/slittamento temporale nello sviluppo dei nervi della mandibola, che a sua volta può essere la conseguenza di un alterato processo di sviluppo dell'osso mandibolare. La sottoespressione di diversi marcatori di sviluppo neuronale sembra suggerire un ritardo nel differenziamento delle cellule che costituiscono la componente nervosa della mandibola.

Ci potrebbe essere comunque una relazione più stretta tra la formazione dell'osso mandibolare e i cambiamenti trascrizionali di geni neurone-specifici, evidenze sempre più numerose indicano infatti che il sistema nervoso partecipa nella regolazione della fisiologia dell'osso (Elefteriou, 2005); neurotrasmettitori periferici esercitano azioni a carico di osteoblasti e osteoclasti che è stato dimostrato esprimono recettori specifici per questi mediatori. Di particolare interesse per questo studio è il ruolo di CRGP, un neuropeptide che agisce a livello del metabolismo osseo promuovendo la formazione dell'osso e reprimendo il suo riassorbimento (Huebner *et al.* 2008); come altri mediatori, CRGP viene anche prodotto direttamente dagli osteoblasti come fattore autocrino (Elefteriou, 2005).

L'espressione di geni neurone-specifici nelle cellule ossee non è limitato ai neurotrasmettitori e i loro recettori ma si estende alla rete molecolare che regola l'esocitosi del glutammato (per es. SNARE, SNAP-25, *syntaxin*, *synaptophysin*, *syntagmin*) descritta a livello delle terminazioni nervose pre-sinaptiche (Bhangu *et al.* 2008). Il ruolo del glutammato negli osteoblasti ed osteoclasti è piuttosto complesso e gli studi *in vivo* sono tuttora limitati (Skerry, 2008); il *knock-out* condizionale in topi mancanti di alcuni componenti della via di segnale del glutammato ha mostrato una ridotta mineralizzazione e una ossificazione ritardata; diversi geni coinvolti nell'esocitosi e nella via di segnalazione del glutammato sono stati riscontrati tra i geni sottoespressi di questo studio (per es. SNAP-25, *Synaptophysin*, *Synapsin*, *Metabotropic glutamate receptor 8*).

Per supportare assieme le evidenze sopradescritte potrebbe essere formulata l'ipotesi secondo cui nei branzini prognati il processo di formazione dell'osso verrebbe ritardato, rispetto agli esemplari normali, attraverso la sottoespressione di diverse reti di segnale che controllano la formazione/rimodellamento dell'osso direttamente negli osteoblasti e/o a carico delle terminazioni nervose che vanno ad innervare la mandibola. Il ritardo

nei processi di ossificazione dell'osso permetterebbe un prolungato sviluppo dei componenti dell'osso mandibolare con la sua conseguente protrusione. Questa ipotesi tuttavia richiede conferme aggiuntive da ulteriori stadi di sviluppo e con l'uso di metodiche sperimentali diverse e indipendenti.



5. CONCLUSIONI GENERALI

A partire dalla scorsa decade, lo sviluppo di nuove tecnologie nel campo della genomica ha rivoluzionato i moderni approcci della ricerca biologica. L'avvento ed il rapido e costante sviluppo della genomica funzionale offre nuove opportunità per lo studio delle interazioni tra i geni e l'ambiente nonché la determinazione di quale gene contribuisce ad una determinata caratteristica fenotipica, fornendo informazioni sul controllo e la regolazione di processi biologici e fisiologici.

La tecnologia *microarray*, permettendo l'analisi sistematica di interi trascrittomi, ha segnato l'inizio dell'era genomica; e se fino a pochi anni fa veniva applicata solo ad una ristretta cerchia di specie modello, è ora di largo consumo ed utilizzata per molteplici scopi.

In questo studio, due piattaforme *microarray* per lo studio dei profili di espressione genica, sono state sviluppate e validate in due specie di grande interesse per il settore ittico europeo; *S. aurata* e *D. labrax*. Entrambe le piattaforme hanno dimostrato una buona riproducibilità e affidabilità, confermata anche mediante validazione con un metodo indipendente quale la *Real-time* RT-PCR. Il disegno di sonde a oligonucleotidi si è dimostrato effettuabile per quasi tutti i trascritti fruibili e le sonde stesse hanno dimostrato un buon successo di ibridazione, permettendo quindi la miglior rappresentazione possibile dei trascrittomi attualmente disponibili per branzino e orata.

L'utilizzo di piattaforme a oligonucleotidi permette di adattare il *design* del *microarray* anche a specie per le quali la conoscenza del trascrittoma è ancora limitata, permettendo l'applicazione di questa tecnologia potenzialmente a qualunque specie allevata e con investimenti economici contenuti. La flessibilità nel *design* di questo tipo di *microarray*, inoltre, rende possibile l'aggiunta o la sostituzione di sonde man mano che nuovi trascritti vengono identificati o nel caso in cui si abbia la necessità di eseguire studi mirati a particolari categorie di geni, adattandone quindi l'applicabilità a seconda delle diverse esigenze.

La conoscenza solo parziale del trascrittoma delle specie ittiche, in particolar modo quelle “non modello”, rappresenta tuttavia una forte limitazione nell’interpretazione dei dati ottenibili mediante le nuove tecnologie della genomica. Se per la maggior parte dei trascritti dei mammiferi (per es. uomo e topo) ed di alcuni invertebrati (*Drosophila melanogaster*) le informazioni sulla loro funzione e regolazione sono velocemente ricavabili; un’adeguata classificazione funzionale dei trascritti in possesso per molte specie ittiche non è, invece, facilmente ottenibile.

Questo problema può essere appianato, anche se non completamente, utilizzando un approccio di tipo comparativo tra le specie “non-modello” e quelle meglio caratterizzate. La classificazione di ciascun gene sulla base dell’omologia di sequenza con genomi/trascrittomi di riferimento permette di attribuire una funzione biologica a ciascun trascritto mediante l’identificazione di “ortologi putativi” piuttosto che sulla base di evidenze sperimentali dirette. L’approccio bioinformatico si presenta, al momento, lo strumento più veloce ed efficace per sopperire, almeno in parte, alle limitazioni dovute alla scarsa conoscenza delle specie “non-modello”.

Nel branzino, la piattaforma *microarray* è stata applicata allo scopo di identificare i geni potenzialmente coinvolti nello sviluppo del prognatismo, deformità che notoriamente colpisce le produzioni di questa specie. Nonostante le limitazioni appena descritte, l’analisi dei risultati ottenuti ha permesso di identificare oltre 200 trascritti la cui espressione è risultata essere significativamente alterata nei soggetti affetti da prognatismo gettando le basi per ulteriori studi e potenzialmente l’identificazione dei processi che determinano sviluppo di questa deformità.

L'analisi approfondita delle componenti dell'immunità innata dei teleostei, comprensiva di relazioni e funzione, offre la prospettiva di generare nuove tecnologie applicabili al miglioramento dell'acquacoltura. Diverse tecniche di biologia molecolare, fornendo gli strumenti per la sua caratterizzazione, hanno aperto la strada ad una migliore conoscenza del sistema immunitario innato dei teleostei. Nel corso di questo dottorato, analisi di genomica comparativa su sequenze BAC di *D. labrax* hanno permesso l'identificazione di un *cluster* di recettori Ig-correlati identificati finora solo nei teleostei, i *Novel Immune-type Receptor*. Utilizzando metodiche di RT-PCR e RACE, tutti i membri di questo complesso genico sono stati clonati e sequenziati, permettendo di studiarne a fondo la struttura e l'organizzazione genomica. La caratterizzazione di questa famiglia genica nel branzino, unitamente agli studi precedentemente effettuati in altre specie di teleostei, ha permesso di acquisire nuove informazioni sull'evoluzione e la funzione di questi recettori, aumentando quindi la conoscenza e la comprensione di questa complessa famiglia recettoriale.

BIBLIOGRAFIA

Abascal F, Zardoya R, Posada D: **ProtTest: selection of best-fit models of protein evolution.** *Bioinformatics* 2005, **21**: 2104-2105.

Abdel I, Abella E, Lopez-Albors O, Valde P, Norte M, Garcia-Alcazar A: **Abnormalities in the juvenile stage of sea bass (*Dicentrarchus labrax* L.) reared at different temperatures: types, prevalence and effect on growth.** *Aquac Intern* 2004, **12**:523-538.

Abzhanov A, Rodda SJ, McMahon AP, Tabin CJ: **Regulation of skeletogenic differentiation in cranial dermal bone.** *Development* 2007, **134**:3133-3144.

Adzhubei AA, Vlasova AV, Hagen-Larsen H, Ruden TA, Laerdahl JK, Høyheim B: **Annotated expressed sequence tags (ESTs) from pre-smolt Atlantic salmon (*Salmo salar*) in a searchable data resource.** *BMC genomics* 2007, **8**:209

Agbayani E (1999): www.fishbase.org/Reproduction/SpawningList, marzo 2009.

Alsina M, Blanch A: **First isolation of *Flexibacter maritimus* from cultivated turbot (*Scophthalmus maximus*).** *Bulletin European Association Fish Pathology* 1993, **13**:157-160.

Anisimova M, Kosiol C: **Investigating protein-coding sequence evolution with probabilistic codon substitution models.** *Mol Biol Evol* 2009, **26**: 255-271.

Aoki T, Takano T, Santos MD, Kondo H and Hirono I: **Molecular Innate Immunity in Teleost Fish: Review and Future Perspectives Fisheries for Global Welfare and Environment.** 5th World Fisheries Congress 2008, 263–276.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.

Bardon A, Vandeputte M, Dupont-Nivet M, Chavanne H, Haffray P, Vergnet A, Chatain B: **What is the heritable component of spinal deformities in the European sea bass (*Dicentrarchus labrax*)?** *Aquaculture* 2009, **294**:194-201.

Bhangu PS, Genever PG, Spencer GJ, Grewal TS, Skerry TM. **Evidence for targeted vesicular glutamate exocytosis in osteoblasts.** *Bone* 2001, **29**:16-23.

Bléry M, Olcese L, Vivier E. **Early signaling via inhibitory and activating NK receptors.** *Hum Immunol* 2000, **61**:51-64

Boglione C., Gagliardi F., Scardi M., Cataudella, S. (2001): **Skeletal descriptors and quality assessment in larvae and post-larvae of wild-caught and hatchery-reared gilthead sea bream (*Sparus aurata* L. 1758).** *Aquaculture*, **192**: 1-22.

Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.** *Bioinformatics* 2003, **19**: 185-193.

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT: **Natural selection on protein-coding genes in the human genome.** *Nature* 2005, **437**: 1153–1157.

Calloni GW, Le Douarin NM, Dupin E: **High frequency of cephalic neural crest cells shows coexistence of neurogenic, melanogenic, and osteogenic differentiation capacities.** *Proc Natl Acad Sci U S A* 2009, **106**:8947-8952.

Campanella JJ, Bitincka L, Smalley J: **MatGAT: An application that generates similarity/identity matrices using protein or DNA sequences.** *BMC Bioinformatics* 2003 **4**:29.

Canavez F, Young NT, Guethlein LA, Rajalingam R, Khakoo SI, Shum BP, Parham P: **Comparison of chimpanzee and human leukocyte Ig-like receptor genes reveals framework and rapidly evolving genes.** *Journ. of Immunol* 2001, **167**: 5786-5794.

Cannon JP, Haire RN, Magis AT, Eason DD, Winfrey KN, Hernandez Prada JA, Bailey KM, Jakoncic J, Litman GW, Ostrov DA: **A bony fish immunological receptor of the NITR multigene family mediates allogeneic recognition.** *Immunity* 2008, **29**: 228–237.

Carninci P: **Constructing the landscape of the mammalian transcriptome.** *Exp Biol* 2007, **210**:1497-1506.

Cavileer T, Hunter S, Okutsu T, Yoshizaki G, Nagler JJ: **Identification of Novel Genes Associated with Molecular Sex Differentiation in the Embryonic Gonads of Rainbow Trout (*Oncorhynchus mykiss*).** *Sex Dev* 2009, **3**:214–224.

Cepeda C, Santos Y: **First isolation of *Flexibacter maritimus* from farmed Senegalese sole (*Solea senegalensis*) in Spain,** *Bulletin European Association Fish Pathology* 2002, **22**:388-392.

Cerdà J, Mercadé J, Lozano JJ, Manchado M, Tingaud-Sequeira A, Astola A, Infante C, Halm S, Viñas J, Castellana B, Asensio E, Cañavate P, Martínez-Rodríguez G, Piferrer F, Planas JV, Prat F, Yúfera M, Durany O, Subirada F, Rosell E, Maes T. **Genomic resources for a commercial flatfish, the Senegalese sole (*Solea senegalensis*): EST sequencing, oligo microarray design, and development of the Soleamold bioinformatic platform.** *BMC Genomics* 2008, **9**:508.

Ceschia G, Quaglio F, Giorgetti G, Bertela G, Bovo G: **Serious outbreak of pasteurellosis (*Pasteurella piscicida*) in euryhaline species along the Italian coasts.** 5th International Conference of the European Association of Fish Pathologists, Budapest, abstract, p. 26, 1991.

Chauvigné F, Cauty C, Rallièrè C, Rescan PY: **Muscle Fiber Differentiation in Fish Embryos as Shown by In Situ Hybridization of a Large Repertoire of Muscle-Specific Transcripts.** *Dev Dyn* 2005, **233**:659-666.

Chavanne H, Chatain B, Haffray P: **Review on Breeding and Reproduction of European aquaculture species.** *Aqua Breeding* 2009, www.aquabreeding.eu.

Chistiakov DA, Hellemans B, Volckaert FA: **Review on the immunology of European sea bass *Dicentrarchus labrax*.** *Vet Immunol Immunopathol* 2007, **117**:1-16

Chistiakov DA, Hellemans B, Haley CS, Law AS, Tsigenopoulos CS, Kotoulas G, Bertotto D, Libertini A, Volckaert FA: **A microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L.** *Genetics* 2005 **170**(4):1821-1826

Chistiakov DA, Tsigenopoulos CS, Lagnel J, Guo YM, Hellemans B, Haley CS, Volckaert FA, Kotoulas G. **A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L.** *Anim Genet* 2008, **39**:623-34.

Chou HH, Holmes MH: **DNA sequence quality trimming and vector removal.** *Bioinformatics* 2001, **17**: 1093-1104.

Chudley AE: **Genetic landmarks through philately - The Habsburg jaw.** *Clin Genet* 1998, **54**:283-284.

Clark MS, Smith SF, Elgar G: **Use of the Japanese pufferfish (*Fugu rubripes*) in comparative genomics.** *Mar Biotechnol* 2001, **3**:S130-40.

Crollius HR: **The tetraodon genome.** *Genome Dyn* 2006, **2**:154-64

Crosetti D, Sola L, Allegrucci G, Libertini A, **Genetica in acquacoltura. Acquacoltura responsabile**, S. Cataudella e P. Bronzi (editori), Unimar-Uniprom2001, Roma.

Cruz RM, Krieger H, Ferreira R, Mah J, Hartsfield J Jr, Oliveira S: **Major gene and multifactorial inheritance of mandibular prognathism**. *Am J Med Genet A* 2008, **146**:71-77.

Desai S, Heffelfinger AK, Orcutt TM, Litman GW, Yoder JA: **The medaka novel immune-type receptor (NITR) gene clusters reveal an extraordinary degree of divergence in variable domains**. *BMC Evolutionary Biology* 2008 **8**:177.

Doron-Faigenboim A, Pupko T: **A combined empirical and mechanistic codon model**. *Mol Biol Evol* 2006, **24**: 388–397.

Douglas SE, Knickle LC, Kimball J, Reith ME: **Comprehensive EST analysis of Atlantic halibut (*Hippoglossus hippoglossus*), a commercially relevant aquaculture species**. *BMC genomics* 2007, **8**:144.

Duret L, Mouchiroud D: **Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate**. *Mol Biol Evol* 2000, **17**:68-74.

EFSA: **Scientific Opinion of the Panel on Biological Hazards on a request from The European Commission on Food Safety considerations concerning the species-specific welfare aspects of the main systems of stunning and killing of farmed fish**. *EFSA Journal* 2009, **1190**:1-16.

Eleftheriou F: **Neuronal signaling and the regulation of bone remodeling**. *Cell Mol Life Sci* 2005, **62**:2339–2349.

Ellegren H: **Sequencing goes 454 and takes large-scale genomics into the wild**. *Mol Ecol* 2008, **17**:1629-1631.

Emrich R, Brodie A, Blayney J: **Prevalence of class I, class II, and class III malocclusions (Angle) in an urban population; an epidemiological study.** *Journal Dental Research*, **44**: 947-953

Emrich SJ, Barbazuk WB, Li L, Schnable PS: **Gene discovery and annotation using LCM-454 transcriptome sequencing.** *Genome Res* 2007, **17**:69-73.

Evenhuis J, Bengtén E, Snell C, Quiniou SM, Miller NW, Wilson M: **Characterization of additional novel immune type receptors in channel catfish, *Ictalurus punctatus*.** *Immunogenetics* 2007, **59**: 661–671

Ewing B, Green P: **Base-calling of automated sequencer traces using phred. II. Error probabilities.** *Genome Res* 1998, **8**: 186-194.

Felsenstein J: *Inferring Phylogenies*. Sinauer, Sunderland, MA 2004.

Ferraresso S, Kuhl H, Milan M, Ritchie DW, Secombes CJ, Reinhardt R, Bargelloni L: **Identification and characterisation of a novel immune-type receptor (NITR) gene cluster in the European sea bass, *Dicentrarchus labrax*, reveals recurrent gene expansion and diversification by positive selection.** *Immunogenetics* 2009, **61**:773-88.

Ferraresso S, Vitulo N, Mininni AN, Romualdi C, Cardazzo B, Negrisol E, Reinhardt R, Canario AVM, Patarnello T, Bargelloni L: **Development and validation of a gene expression oligo microarray for the gilthead sea bream (*Sparus aurata*).** *BMC Genomics* 2008, **9**:580.

Food and Agricultural Organisation, 2008. **The State of World Fisheries and Aquaculture.** Editorial Group FAO Information Division.

Franch R, Louro B, Tsalavouta M, Chatziplis D, Tsigenopoulos CS, Sarropoulou E, Antonello J, Magoulas A, Mylonas CC, Babbucci M, Patarnello T, Power DM, Kotoulas G, Bargelloni L. **A genetic linkage map of the hermaphrodite teleost fish *Sparus aurata* L.** *Genetics* 2006, 174:851-61.

Gauthier G, Lafay B, Ruimy R, Breittmayer V, Nicolas J, Gauthier M, Christen R: **Small subunit rRNA sequences and whole DNA relatedness concur for the reassignment of *Pasteurella piscicida* to the genus *Photobacterium* as *Photobacterium damsela* subsp. *piscicida*,** *Internat Journ of System Bacteriol* 1995, 45:139-144.

Geay F, Darias MJ, Santigosa E, Desbruyères E, Quazuguel P, Zambonino-Infante JL, Cahu CL, Mazurais D: **Cloning of endothelin-1 (ET-1) from European sea bass (*Dicentrarchus labrax*) and its gene expression analysis in larvae with retinoic acid-induced malformations.** *Aquaculture* 2009, 287:169–173.

Geoghegan F, Katsiadaki I, Williams TD, Chipman JK: **A cDNA microarray for the three-spined stickleback, *Gasterosteus aculeatus* L., and analysis of the interactive effects of oestradiol and dibenzanthracene exposures.** *J Fish Biol* 2008, 72:2133-2153.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO suite.** *Nucleic Acids Res* 2008, 36:3420-35.

Govoroun M, Le Gac F, Guiguen Y: **Generation of a large scale repertoire of Expressed Sequence Tags (ESTs) from normalised rainbow trout cDNA libraries.** *BMC Genomics* 2006, 7:196.

Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, 52: 696-704

Hacia JG: Resequencing and mutational analysis using oligonucleotide microarrays. *Nat Genet* 1999, **21**:42-7.

Haga Y, Suzuki T, Takeuchi T: **Retinoic acid isomers produce malformations in postembryonic development of the Japanese flounder, *Paralichthys olivaceus*.** *Zoolog Sci* 2002, **10**:1105-12.

Hall TE, Cole NJ, Johnston IA: **Temperature and the expression of seven muscle-specific protein genes during embryogenesis in the Atlantic cod *Gadus morhua* L.** *J Exp Biol* 2003, **206**:3187-3200.

Hawke J, Plakas S, Minton R, McPhearson R, Snider R, Guarino M: **Fish Pasteurellosis of cultured striped bass (*Morone saxatilis*) in coastal Alabama.** *Aquaculture* 1997, **65**:193-204.

Hawke NA, Yoder JA, Haire RN, Mueller MG, Litman RT, Miracle AL, Stuge T, Shen L, Miller N, Litman GW: **Extraordinary variation in a diversified family of immune-type receptor genes.** *Proc Natl Acad Sci U S A.* 2001, **98**: 13832–13837

Hu ZL, Bao J, Reecy JM: CateGORizer: **A web-based program to batch analyze gene ontology classification categories.** *Online J Bioinform* 2008, **9**:108-112

Huang CM, Mi MP, Vogt DW: **Mandibular prognathism in the rabbit: discrimination between single-locus and multifactorial models of inheritance.** *J Hered* 1981, **72**:296-298.

Huang X, Madam A: **CAP3: a DNA sequence assembly program.** *Genome Res* 1999, **9**:868-877.

Huebner AK, Keller J, Catala-Lehnen P, Perkovic S, Streichert T, Emeson RB, Amling M, Schinke T: **The role of calcitonin and alpha-calcitonin gene-related peptide in bone formation.** *Arch Biochem Biophys* 2008, **473**:210-217.

Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephanians SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, Linsley PS: **Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer.** *Nat Biotechnol* 2001, **19**:342-7.

Imai S, Heino TJ, Hienola A, Kurata K, Büki K, Matsusue Y, Väänänen HK, Rauvala H: **Osteocyte-derived HB-GAM (pleiotrophin) is associated with bone formation and mechanical loading.** *Bone* 2009, **44**:785-794.

Teraoka H, Dong W, Okuhara Y, Iwasa H, Shindo A, Hill AJ, Kawakami A, Hiraga T: **Impairment of lower jaw growth in developing zebrafish exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin and reduced hedgehog expression.** *Aquatic Toxicology* 2006, **78**:103-113.

Jacquier A: **The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs.** *Nat Rev Genet* 2009, **10**:833-844

Jezewski PA, Fang PK, Payne-Ferreira TL, Yelick PC: **Alternative splicing, phylogenetic analysis, and craniofacial expression of zebrafish *tbx22*.** *Dev Dyn* 2009, **238**:1605-1612.

Jobb G, von Haeseler A, Strimmer K: **TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics.** *BMC Evol Biol* 2004: 4, 18.

Jørgensen SM, Afanasyey S, Krasnov A: **Gene expression analyses in Atlantic salmon challenged with infectious salmon anemia virus reveal differences between individuals with early, intermediate and late mortality.** *BMC Genomics* 2008, **9**:179.

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, Jindo T, Kobayashi D, Shimada A, Toyoda A, Kuroki Y, Fujiyama A, Sasaki T, Shimizu A, Asakawa S, Shimizu N, Hashimoto S, Yang J, Lee Y, Matsushima K, Sugano S, Sakaizumi M, Narita T, Ohishi K, Haga S, Ohta F, Nomoto H, Nogata K, Morishita T, Endo T, Shin-I T, Takeda H, Morishita S, Kohara Y: **The medaka draft genome and insights into vertebrate genome evolution.** *Nature* 2007, **447**:714-719.

Kassahn KS, Dang VT, Wilkins SJ, Perkins AC, Ragan MA: **Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates.** *Genome Res* 2009, **19**:1404-18.

Kawahara E, Kusuda R: **Location of Pasteurella piscicida antigens in tissues of yellowtail Seriola quinqueradiata vaccinated by immersion.** *Bullet of the Jap Soc of Scient fisher* 1997, **54**:1101-1105.

Klaper R, Carter BJ, Richter CA, Drevnick PE, Sandheinrich MB, Tillitt DE: **Use of a 15 k gene microarray to determine gene expression changes in response to acute and chronic methylmercury exposure in the fathead minnow *Pimephales promelas* Rafinesque.** *J Fish Biol* 2008, **72**:2207-2280.

Kock H, Fischer U. **A novel immunoglobulin-like transcript from rainbow trout with two Ig-like domains and two isoforms.** *Mol Immunol.* 2008, **45(6)**: 1612-1622.

Kosakovsky Pond SL, Frost SD: **Not so different after all: a comparison of methods for detecting amino acid sites under selection.** *Mol Biol Evol* 2005, **22**: 1208-1222.

Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A: **Patterns of positive selection in six Mammalian genomes.** *PLoS Genet* 2008, **4**: e1000144.

Kuzir S, Kozaric S, Nejedli S (2004): **Development of mandibular arch in European sea bass, *Dicentrarchus labrax* from the “Cenmar” hatchery, Croatia.** *Veterinarski arhiv*, **74**:321-330.

Kuzniar A, van Ham RC, Pongor S, Leunissen JA. **The quest for orthologs: finding the corresponding gene across genomes.** *Trends Genet* 2008, **24**:539-551.

Leder EH, Merilä J, Primmer CR: **A flexible whole-genome microarray for transcriptomics in three-spine stickleback (*Gasterosteus aculeatus*).** *BMC Genomics* 2009, **10**:426.

Lefranc MP, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Troung L, Thouvenin-Contet V, Lefranc G: **IMGT unique numbering for immunoglobulin an T cell receptor variable domains and Ig superfamily V-like domains.** *Dev Comp Immunol* 2003, **27**: 55-77

Lehner B, Fraser AG: **Protein domains enriched in mammalian tissue-specific or widely expressed genes.** *TRENDS Genet* 2004, **20**:468-472.

Levi L, Pekarski I, Gutman E, Fortina P, Hyslop T, Biran J, Levavi-Sivan B and Lubzens E: **Revealing genes associated with vitellogenesis in the liver of the zebrafish (*Danio rerio*) by transcriptome profiling.** *BMC Genomics* 2009, **10**:141.

Lewis SE, Searle SMJ, Harris N, Gibson M, Iyer V, Richter J, Wiel C, Bayraktaroglu L, Birney E, Crosby MA, Kaminker JS, Matthews BB, Prochnik SE, Smith CD, Tupy JL, Rubin GM, Misra S, Mungall CJ, Clamp ME: **Apollo: a sequence annotation editor.** *Genome Biol* **3**: research0082.

Li P, Peatman E, Wang S, Feng J, He C, Baoprasertkul P, Xu P, Kucuktas H, Nandi S, Somridhivej B, Serapion J, Simmons M, Turan C, Liu L, Muir W, Dunham R, Brady Y, Grizzle J, Liu Z: **Towards the ictalurid catfish transcriptome: generation and analysis of 31,215 catfish ESTs.** *BMC Genomics* 2007, **8**:177.

- Li RW, Waldbieser GC: **Production and utilization of a high-density oligonucleotide microarray in channel catfish, *Ictalurus punctatus*.** *BMC Genomics* 2006, **7**:134.
- Litman GW, Hawke NA, Yoder JA: **Novel immune-type receptor genes.** *Immunological Reviews* 2001, **181**: 250-259.
- Liu R, Li RW, Waldbieser GC: **Utilization of microarray technology for functional genomics in ictalurid catfish.** *J Fish Biol* 2008, **72**:2377-2390.
- Long EO: **Negative signaling by inhibitory receptors: the NK cell paradigm.** *Immunol Rev* 2008, **224**: 70–84
- MAQC Consortium, Shi L, Reid LH, Jones WD, Shippy R, Warrington JA et al. .: **The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.** *NatBiotechnol* 2006, **24**:1151-1161.
- Martin AM, Freitas EM, Witt CS, Christiansen FT: **The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster.** *Immunogenetics* 2000, **51**: 268–280.
- Martin SAM, Blaney SC, Houlihan DF, Secombes CJ: **Transcriptome response following administration of a live bacterial vaccine in Atlantic salmon (*Salmo salar*).** *Mol Immunol* 2006, **43**:1900-1911.
- Martinez-Borra J, Khakoo SI: **Speed and selection in the evolution of killer-cell immunoglobulin-like receptors.** *Int J Immunogenet* 2008, **35**: 89–96.
- Mazurais D, Glynatsi N, Darias M, Christodouloupolou S, Cahu C, Zambonino-Infante J-L, Koumoundouros G: **Optimal levels of dietary vitamin A for reduced deformity incidence during development of European sea bass larvae (*Dicentrarchus labrax*) depend on malformation type.** *Aquaculture* 2009, **294**:262–270.

Meiri I, Gothilf Y, Zohar Y, and Elizur A: **Physiological changes in the spawning gilthead seabream, Sparus aurata, succeeding the removal of males.** *Journ of Experiment Zool* 2002, **292**:555–564.

Melotti P, Roncarati A: **State of the art and future trends of European and Italian aquaculture.** *Vet Res Commun* 2009 33. **Suppl 1**:S9–S13

Meyaard L, Adema GJ, Chang C, Woollatt E, Sutherland GR, Lanier LL, Phillips JH: **LAIR-1, a novel inhibitory receptor expressed on human mononuclear leukocytes.** *Immunity* 1997, **7**: 283-290.

Millán A, Gómez-Tato A, Fernández C, Pardo GB, Álvarez-Dios JA, Calaza M, Bouza C, Vázquez M, Cabaleiro S, Martínez P: **Design and Performance of a Turbot (Scophthalmus maximus) Oligo-microarray Based on ESTs from Immune Tissues.** *Mar Biotechnol* 2009 Oct 21.

Miller CT, Yelon D, Stainier DY, Kimmel CB: **Two endothelin 1 effectors, hand2 and bapx1, pattern ventral pharyngeal cartilage and the jaw joint.** *Development* 2003, **130**:1353-1365.

Miller KM, Maclean N.: **Teleost microarrays: development in a broad phylogenetic range reflecting diverse applications.** *Journal of Fish Biology* 2008, **72**:2039–2050.

Morey JS, Ryan JC, VanDolah FM: **Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR.** *Biol Proced Online* 2006, **8**:175-193.

Nakatani Y, Takeda H, Kohara Y, Morishita S: **Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates.** *Genome Res* 2007, 17(9): 1254-1265.

Nascimento DS, Pereira PJ, Reis MI, do Vale A, Zou J, Silva MT, Secombes CJ, dos Santos NM. **Molecular cloning and expression analysis of sea bass (Dicentrarchus**

labrax L.) tumor necrosis factor-alpha (TNF-alpha). *Fish Shellfish Immunol.* 2007, **23**:701-10.

Nègre N, Lavrov S, Hennetin J, Bellis M, Cavalli G: **Mapping the distribution of chromatin proteins by ChIP on chip.** *Methods Enzymol* 2006, **410**:316-41.

Nei M, Rooney AP: **Concerted and birth-and-death evolution of multigene families.** *Annu Rev Genet* 2005, **39**: 121-152

Norman PJ, Abi-Rached L, Gendzekhadze K, Korbel D, Gleimer M, Rowley D, Bruno D, Carrington CV, Chandanayingyong D, Chang YH, Crespí C, Saruhan-Direskeneli G, Fraser PA, Hameed K, Kamkamidze G, Koram KA, Layrisse Z, Matamoros N, Milà J, Park MH, Pitchappan RM, Ramdath DD, Shiau MY, Stephens HA, Struik S, Verity DH, Vaughan RW, Tyan D, Davis RW, Riley EM, Ronaghi M, Parham P: **Unusual selection on the KIR3DL1/S1 natural killer cell receptor in Africans.** *Nat Genet* 2007, **39**: 1092-1099.

Oostlander AE, Meijer GA, Ylstra B: **Microarray-based comparative genomic hybridization and its applications in human genetics.** *Clin Genet* 2004, **66**:488-95.

Østergaard AE, Martin SAM, Wang T, Stet RJM, Secombes CJ: **Rainbow trout (*Oncorhynchus mykiss*) possess multiple novel immunoglobulin-like transcripts containing either an ITAM or ITIMs.** *Dev Comp Immunol* **33**: 525-532

Pardo BG, Fernández C, Millán A, Bouza C, Vázquez-López A, Vera M, Alvarez-Dios JA, Calaza M, Gómez-Tato A, Vázquez M, Cabaleiro S, Magariños B, Lemos ML, Leiro JM, Martínez P: **Expressed sequence tags (ESTs) from immune tissues of turbot (*Scophthalmus maximus*) challenged with pathogens.** *BMC Vet Res.* 2008, **4**:37.

Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, Collins PJ, Chu TM, Bao W, Fang H, Kawasaki ES, Hager J, Tikhonova IR, Walker SJ, Zhang L, Hurban P, De Longueville F, Fuscoe JC, Tong W, Shi L, Wolfinger RD: **Performance comparison**

of one-color and two-color platforms within the MicroArray Quality Control (MACQ) project. *Nat Biotechnol* 2006, **24**:1140-1150.

Patthy L: **Exons: Shuffling.** In: **Encyclopedia of Life Sciences (ELS).** John Wiley & Sons. Ltd: Chichester.

Peatman E, Terhune J, Baoprasertkyl P, Xu P, Nandi S, Wang S, Somridhivej B, Kucuktas H, Li P, Dunham R, Liu Z: **Microarray analysis of gene expression in the blue catfish liver reveals early activation of the MHC class I pathway after infection with *Edwardsiella ictaluri*.** *Mol Immunol* 2008, **45**:553-566.

Pepitone C., Massari F., Thomas M., **I pesci delle acque costiere italiane,** L'EPOS 1995.

Piyaviriyakul P, Kondo H, Hirono I, Aoki T: **A novel immune-type receptor of Japanese flounder (*Paralichthys olivaceus*) is expressed in both T and B lymphocytes.** *Fish Shellfish Immunol* 2007, **22**: 467-476.

Pond SL, Frost SD: **Datamonkey: rapid detection of selective pressure on individual sites of codon alignments.** *Bioinformatics* 2005, **21**: 2531-2533.

Porcellotti S (2005): **Pesci d'Italia, ittiofauna delle acque dolci.** Plan, Firenze.

Rajalingam R, Parham P, Abi-Rached L: **Domain shuffling has been the main mechanism forming new hominoid killer cell Ig-like receptors.** *J Immunol* (2004), **172** (1): 356-369.

Randelli E, Buonocore F, Casani D, Fausto AM, Scapigliati G: **An “immunome” gene panel for transcriptomic analysis of immune defence activities in the teleost sea bass (*Dicentrarchus labrax* L.): a review.** *Italian Journal of Zoology* 2009, **76**: 146-157.

Raso M. 1981: **Anatomia patologica clinica,** v. 2 , Piccin 1981, Padova.

Ravagnan G. **Vallicoltura integrata**, Edagricole 1992, Bologna.

Reppe S, Rian E, Jemtland R, Olstad OK, Gautvik VT, Gautvik KM: **Sox-4 messenger RNA is expressed in the embryonic growth plate and regulated via the parathyroid hormone/parathyroid hormone-related protein receptor in osteoblast-like cells.** *J Bone Miner Res* 2000,**15**:2402-2412.

Rescan PY: **New insights into skeletal muscle development and growth in teleost fishes.** *J. Exp. Zool. (Mol. Dev. Evol.)* 2008, **310B**.

Rhee SY, Wood V, Dolinski K, Draghici S: **Use and misuse of the gene ontology annotations.** *Nat Rev Genet* 2008, **9**:509-515.

Ritchie B, Harrison G, Harrison L, Zantop D: **Avian Medicine: Principles and Application** (1997). Abridged Edition.

Robalino J, Carnegie RB, O'Leary N, Ouvry-Patat SA, De la Vega E, Prior S, Gross PS, Browdy CL, Robert W, Kevin C, Schey L, Warr G: **Contributions of functional genomics and proteomics to the study of immune responses in the Pacific white leg shrimp *Litopenaeus vannamei*.** *Vet Immunol Immunopath* 2009, **128**:110–118.

Rosamond L. Naylor¹, Ronald W. Hardy², Dominique P. Bureau³, Alice Chiu⁴, Matthew Elliott⁵, Anthony P. Farrelle⁶, Ian Forstere⁷, Delbert M. Gatlin⁸, Rebecca J. Goldburgh⁹, Katheline Huac¹⁰, and Peter D. Nichols¹¹: **Feeding aquaculture in an era of finite resources:** *Proc Natl Acad Sci U S A* 2009, **16**:15103–15110

Rudolph MG, Wilson IA: **The specificity of TCR/pMHC interaction.** *Curr Opin Immunol.* 2002, **14**:52-65.

Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints.** *J Mol Biol* 1993, **234**: 779-815

Samadiris J, Colonna M: **Cloning of novel immunoglobulin superfamily receptors expressed on human myeloid and lymphoid cells: structural evidence for new stimulatory and inhibitory pathways.** *Eu. J Immunol.* 1997 **27**: 660-665.

Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors.** *Proc Natl Acad Sci U S A* 1977, **74**:5463-7.

Sarropoulou E, Franch R, Louro B, Power DM, Bargelloni L, Magoulas A, Senger F, Tsalavouta M, Patarnello T, Galibert F, Kotoulas G, Geisler R. **A gene-based radiation hybrid map of the gilthead sea bream *Sparus aurata* refines and exploits conserved synteny with *Tetraodon nigroviridis*.** *BMC Genomics* 2007, **7**:8:44.

Sarropoulou E, Kotoulas G, Power DM, Geisler R: **Gene expression profiling of gilthead sea bream during early development and detection of stress-related genes by the application of cDNA microarray technology.** *Physiol Genomics* 2005, **23**:182-191

Sarropoulou E, Sepulcre P, Poisa-Beiro L, Mulero V, Meseguer J, Figueras A, Novoa B, Terzoglou V, Reinhardt R, Magoulas A, Kotoulas G: **Profiling of infection specific mRNA transcripts of the European seabass *Dicentrarchus labrax*.** *BMC Genomics* 2009, **10**:10:157

Scapigliati G, Buonocore F, Randelli E, Casani D, Meloni S, Zarletti G, Tiberi M, Pietretti D, Boschi I, Manchado M, Martin-Antonio B, Jimenez-Cantizano R, Bovo G, Borghesan F, Lorenzen N, Einer-Jensen K, Adams S, Thompson K, Alonso C, Bejar J, Cano I, Borrego JJ, Alvarez MC: **Cellular and molecular immune responses of the sea bass (*Dicentrarchus labrax*) experimentally infected with betanodavirus.** *Fish Shellfish Immunol* 2009, [Epub ahead of print].

Scapigliati G, Romano N, Buonocore F, Picchiatti S, Baldassini MR, Prugnoli D, Galice A, Meloni S, Secombes CJ, Mazzini M, Abelli L: **The immune system of sea bass, *Dicentrarchus labrax*, reared in aquaculture.** *Dev Comp Immunol* 2002, **26**: 151-160.

Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 2005, **270**:467-70.

Senger F, Priat C, Hitte C, Sarropoulou E, Franch R, Geisler R, Bargelloni L, Power D, Galibert F: **The first radiation hybrid map of a perch-like fish: the gilthead seabream (*Sparus aurata* L)** *Genomics* 2006, **87**:793-800.

Sepulcre MP, Sarropoulou E, Kotoulas G, Meseguer J, Mulero. V. ***Vibrio anguillarum* evades the immune response of the bony fish sea bass (*Dicentrarchus labrax* L.) through the inhibition of leukocyte respiratory burst and down-regulation of apoptotic caspases.** *Mol Immunol.* 2007, 44:3751-7

Skerry TM. **The role of glutamate in the regulation of bone mass and architecture.** *J Musculoskelet Neuronal Interact* 2008, **8**(2):166-73.

Sorek R: **The birth of new exons: mechanisms and evolutionary consequences.** *RNA* 2007, 13: 1603-1608.

Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T **Selecton: Advanced models for detecting positive and purifying selection using a Bayesian inference approach.** *Nucleic Acids Res* 2007, **35**:W506-W511.

Stet RJM, Hermsen T, Westphal AH, Jukes J, Engelsma M, Verburg-van Kemenade BML, Dortmans J, Aveiro J, Savelkoul HFJ: **Novel immunoglobulin-like transcripts in teleost fish encode polymorphic receptors with cytoplasmic ITAM or ITIM and a new structural Ig domain similar to the natural cytotoxicity receptor NKp44.** *Immunogenetics.* 2005, **57**: 77–89.

Strong SJ, Mueller MG, Litman RT, Hawke NA, Haire RN, Miracle AL, Rast JP, Amemiya CT, Litman GW: **A novel multigene family encodes diversified variable regions.** *Proc Natl Acad Sci U S A* 1999, **96**: 15080–15085.

Subramaniam S, Kumar S: **Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome.** *Genetics* 2004, **168**:373-381.

Takeda H: **Draft genome of the medaka fish: a comprehensive resource for medaka developmental genetics and vertebrate evolutionary biology.** *Dev Growth Differ.* 2008, **1**:S157-66.

Takeuchi, T, Dedi J, Haga Y, Seikai, T, Watanabe T: **Effect of vitamin A compounds on bone deformity in larval Japanese flounder (*Paralichthys olivaceus*).** *Aquaculture* 1998, **169**: 155– 165.

Teh C, Parinov S, Korzh V: **New ways to admire zebrafish: progress in functional genomics research methodology.** *Biotechniques* 2005, **38**:897-906.

Teraoka H, Donga W, Okuhara Y, Iwasa H, Shindo A, Hill A, Kawakami A, Hiraga T:

Thompson JD, Higgins DG, Gibson TJ CLUSTAL W: **Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22(22)**: 4673-4680.

Tingaud-Sequeira A, Chauvigné F, Lozano J, Agulleiro AM, Asensio E, Cerdà J: **New insights into molecular pathways associated with flatfish ovarian development and atresia revealed by transcriptional analysis.** *BMC Genomics* 2009, **10**:434.

Torkar M, Norgate Z, Colonna M, Trowsdale J, Wilson MJ: **Isotypic variation of novel Immunoglobulin-like transcript/killer cell inhibitory receptor loci in the leukocyte receptor complex.** *Eur J Immunol.* 1998, **28**: 3959–3967.

Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98**:5116-5121.

Villeneuve DL, Knoebl I, Larkin P, Miracle AL, Carter BJ, Denslow ND, Ankley GT: **Altered gene expression in the brain and liver of female fathead minnows *Pimephales promelas* Rafinesque exposed to fadrozole.** *J Fish Biol* 2008, **72**:2281-2340.

Villeneuve L., Gisbert E, Zambonino-Infante JL: **Effect of nature of dietary lipids on European sea bass morphogenesis: implication of retinoid receptors.** *Brit J Nut* 2005, **94**:877-884.

Von Schalburg KR, Cooper GA, Leong J, Robb A, Lieph R, Rise ML, Davidson WS, Koop BF: **Expansion of the genomics research on Atlantic salmon *Salmo salar* L. project (GRASP) microarray tools.** *J Fish Biol* 2008, **72**:2051-2070.

Wang Y, Barbacioru C, Hyland F, Xiao W, Hunkapiller KL, Blake J, Chan F, Gonzalez C, Zhang L, Samaha RR: **Large scale real-time PCR validation on gene expression measurements from two commercial long-oligonucleotide microarrays.** *BMC genomics* 2006, **7**:59.

Wei S, Zhou JM, Chen X, Shah RN, Liu J, Orcutt TM, Traver D, Djeu JY, Litman GW, Yoder JA **The zebrafish activating immune receptor Nitr9 signals via Dap12.** *Immunogenetics* 2007, **59**: 813-821.

Whitaker HA, McAndrew BJ, Taggart JB. **Construction and characterization of a BAC library for the European sea bass *Dicentrarchus labrax*.** *Anim Genet* 2006, **37**:526.

Wolber PK, Collins PJ, Lucas AB, De Witte A, and Shannon KW: **The Agilent in situ-synthesized microarray platform.** *Methods in enzymology* 2006 **410**:28-57.

Wolff G, Wienker T, Sander H: **On the genetics of mandibular prognathism: analysis of large European noble families.** *Journal of Medical Genetics* 1993, **30**:112-116.

Wolfsberg TG, Madden TL: **Sequence similarity searching using the BLAST family of programs.** *Curr Protoc Mol Biol* 2001, 19 Unit 19.3.

Xu Y, He J, Wang X, Lim TM, Gong Z: **Asynchronous Activation of 10 Muscle-Specific Protein (MSP) Genes During Zebrafish Somitogenesis.** *Dev Dyn* 2000, **219**(2): 201-215.

Yamaguchi T: **Genome-wide Linkage Analysis of Mandibular Prognathism in Korean and Japanese Patients.** *Journal Dental Research* 2005, **84**:255-259.

Yang Z, Wong WS, Nielsen R. **Bayes empirical bayes inference of amino acid sites under positive selection.** *Mol Biol Evol* 2005, **22**: 1107-1118.

Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, 24(8): 1586-1591.

Yin JQ, Zhao RC, Morris KV: **Profiling microRNA expression with microarrays.** *Trends Biotechnol* 2008, **26**:70-6.

Yoder JA, Cannon JP, Litman RT, Murphy C, Freeman JL, Litman GW: **Evidence for a transposition event in a second NITR gene cluster in zebrafish.** *Immunogenetics* 2008, **60**: 257–265

Yoder JA, Litman RT, Mueller MG, Desai S, Dobrinski KP, Montgomery JS, Buzzeo MP, Ota T, Amemiya CT, Trede NS, Wei S, Djeu JY, Humphray S, Jekosch K, Hernandez Prada JA, Ostrov DA, Litman GW: **Resolution of the novel immune-type receptor gene cluster in zebrafish.** *Proc Natl Acad Sci U S A* 2004, **101**: 15706–15711.

Yoder JA, Mueller MG, Nichols KM, Ristow SS, Thorgaard GH, Ota T, Litman GW: **Cloning novel immune-type inhibitory receptors from the rainbow trout, *Oncorhynchus mykiss*.** *Immunogenetics* **54**: 662–670.

Yoder JA, Mueller MG, Wei S, Corliss BC, Prather DM, Willis T, Litman RT, Djeu JY, Litman GW: **Immune type receptor genes in zebrafish share genetic and functional properties with genes encoded by the mammalian lymphocyte receptor cluster.** *Proc Natl Acad Sci USA* 2001, **98**: 6771–6776.

Yoder JA, Orcutt TM, Traver D, Litman GW: **Structural characteristics of zebrafish orthologs of adaptor molecules that associate with transmembrane immune receptors.** *Gene* 2007, **401**: 154–164.

Yoder JA: **Form, function and phylogenetics of NITRs in bony fish.** *Dev Comp Immunol* 2009, **33**: 135-144.

Yokoyama WM, Plougastel BFM: **Immune functions encoded by the natural killer gene complex.** *Nat Rev Immunol*, **3**: 304-316

Zhang L, Li WH: **Mammalian housekeeping genes evolve more slowly than tissue-specific genes.** *Mol Biol Evol* 2004, **21**:236-239.

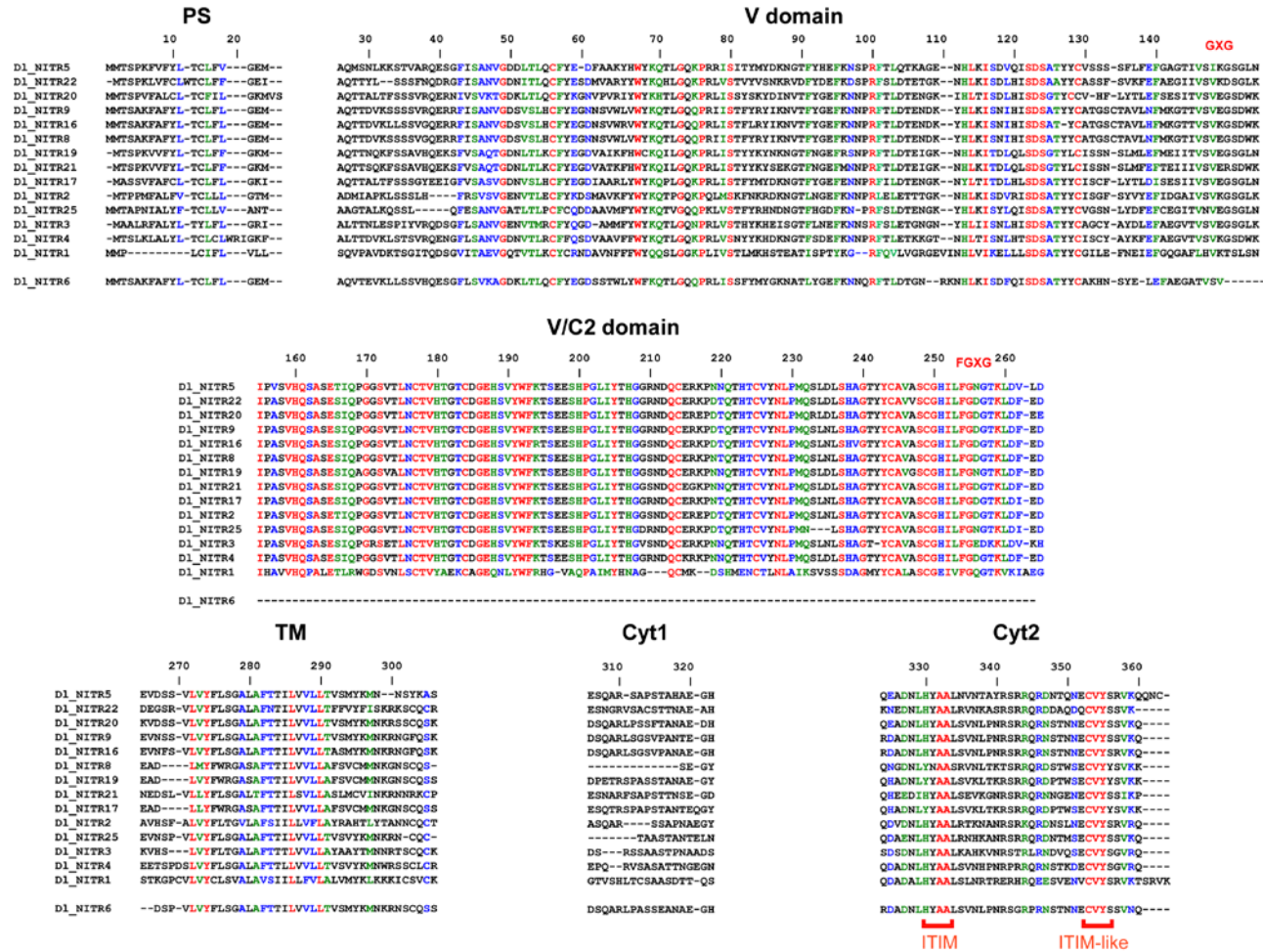
APPENDICI

APPENDICE 2A

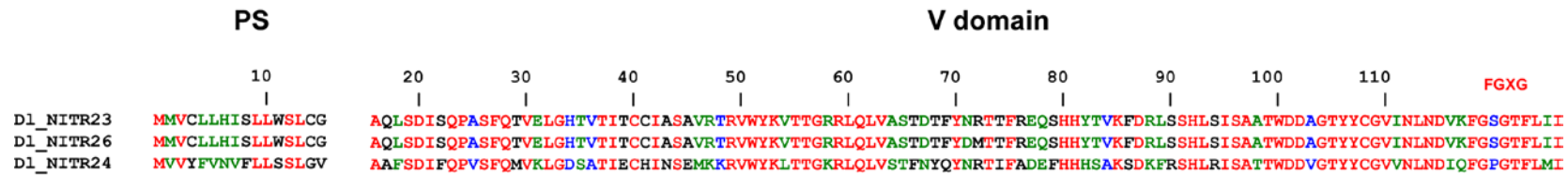
Coppie di *primer* gene-specifiche utilizzate per l'amplificazione dei 27 geni NITR identificati in branzino

Nome	Seq. Primer FOR (5'-3')	Seq Primer REV (3'-5')	T° (*)
NITR1	TGTCGGAATGATGCCGTAA	TTCTCTGGTTCTGTTTCAGACTCAG	62°
NITR2	GCTGATGTCTAAATTCAATAAGCG	TCTGTTGGCATTTCCTGGTCC	59°
NITR3	GCCCATATATGTGCGTCAAGAT	ACCTTGTGTGCCTTCAAAGC	61°
NITR4	TTGCAGCAGTGTTCCTCTGG	CACATTCATCCTTGGTGCTG	62°
NITR5	AAAAGTCCACGGTGGCG	GATCTGTAGGCGGTGTTAACATT	61°
NITR6	AAGCAGGAGACAAGTTGACCC	CATTTGCCTCGGAGGAAG	62°
NITR8	TGCTTTCTATCTGACATGTTTGTTTC	CACTCCAGGTTGAATCCCG	62°
NITR9	AAGGGCACTACTGTCTGAGTGTG	ACGACCATCCATGACTATCCA	60°
NITR10	GCCTGTGGTTTCCTTGTGGT	TTGAAATGAGCTCCAACAACCTC	61°
NITR11	TCAGGTCTACTGACCCTGCTAAAT	ACACGAGTCTCCACTTGTCTCTC	60°
NITR12	CCAACAGACCAAACGCCA	GCTGGTTCTCTGATGGAGAACTT	60°
NITR13	GCTGAACTCTGATGCCAATGT	TCAGGTGATCAGAGCCCAG	60°
NITR14	TACTTGTTCCTGTTTCAGGCTG	AGTTTCCACCTGACTGAAATCC	60°
NITR15	CCAGCAAATTTGAATCTGACTCT	TTCCCTCTTTTGTAAAGCTCCAC	60°
NITR16	TTCAGACTCAGCTACATACTGCG	CCATCCACCCCCAAGAA	60°
NITR17	TTACTTGACAATCACAGATTTGCA	GATCTCTTTGTGCTGTTGACGCT	62°
NITR19	TTCACCTAATGTTGGAGTTTATGGAA	TGATCTCTTCGTCAGTTTGACAC	62°
NITR20	TTCAGTTAAACCAGGGGACAA	GGTTGTCTGCCTCTGGTGAT	59°
NITR21	GTGAAAAGGGAACTTTTAATGGTG	ATCTGTTGCCCTTGAAGTTCG	62°
NITR22	GACAGTCCGCGTTTTTCC	GTGCGTCGTCCTCTGTTC	60°
NITR23	AGCATATCTGCAGCAACGTG	TCTCCCTTAGAGCACCGTGT	62°
NITR24	AAATGGTGAAGCTGGGTGAC	GAAACGTCTCCGTTCTGTTGA	59°
NITR25	CCGTGCTTCTGTCAAGATGA	CTTATGATTCTTAAAGCAGCG	59°
NITR26	CAACTGGTGGCTTCAACAGATA	CAGGAGGAGCCTCTAGGACC	60°
NITR28	AGTTACAATTCAGGCGGAGTG	ACATGTGGCTCCAGCTCTCT	60°
NITR29	GTCTGTACAGTGTGAAGGGA	TTAACAGCCAGAACAATGAGA	60°
NITR30	TCTATGTTTGGATCTGATGGC	GATTTAACAGTCAGAACAATGACA	60°

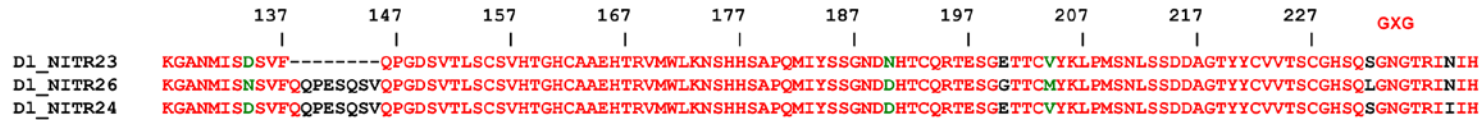
(*) Temperatura di annealing di ciascuna coppia di primer.



Allineamento aminoacidico dei geni NITR appartenenti al Gruppo I. I residui identici sono evidenziati in rosso, i residui conservati ei poco conservati sono evidenziati rispettivamente in verde e blu. Sono rappresentati anche i motivi *Joining* nonché le sequenze ITIM o *ITIM-like*.



VIC2 domain



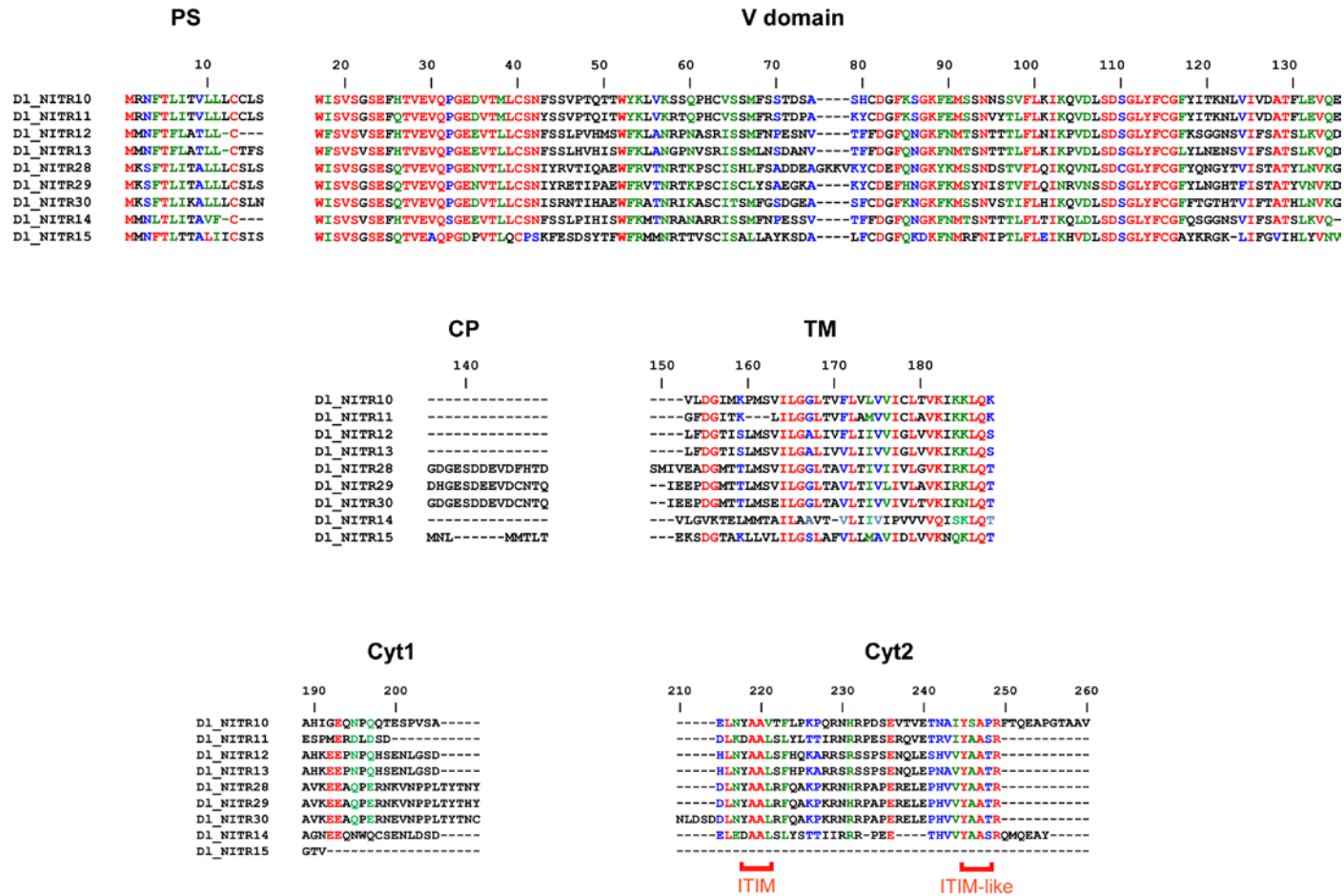
TM

Cyt1

Cyt2

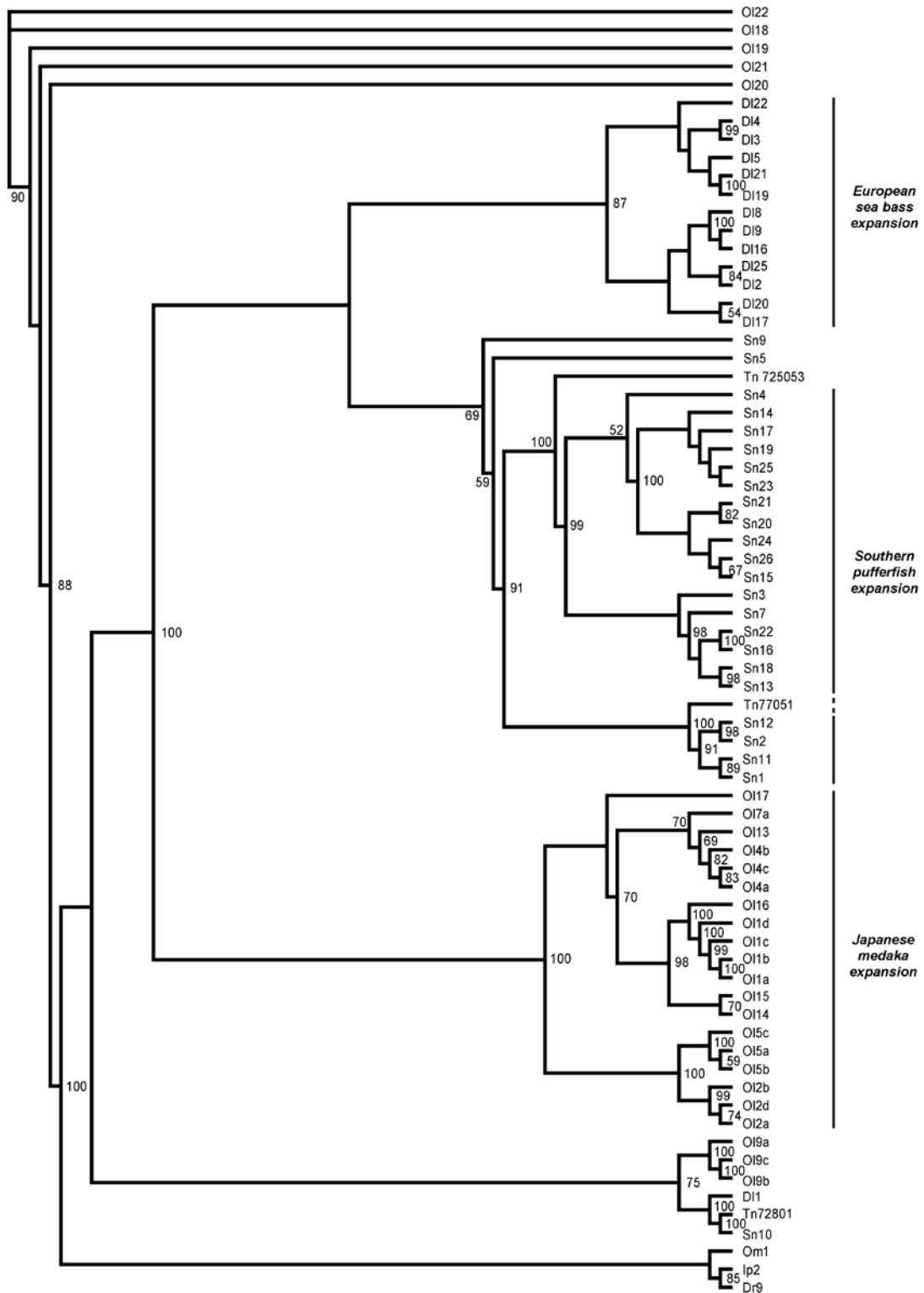


Allineamento aminoacidico dei geni NITR appartenenti al **Gruppo II**. I residui identici sono evidenziati in rosso, i residui conservati ei poco conservati sono evidenziati rispettivamente in verde e blu. Sono rappresentati anche i motivi *Joining* nonché le sequenze ITIM o *ITIM-like*.



Allineamento aminoacidico dei geni NITR appartenenti al **Gruppo III**. I residui identici sono evidenziati in rosso, i residui conservati ei poco conservati sono evidenziati rispettivamente in verde e blu. Sono rappresentate anche le sequenze ITIM o *ITIM-like*.

APPENDICE 2C



Albero filogenetico che mostra le relazioni evolutive tra i domini V e I. Le sequenze pubbliche dei NITR di *S. nephelus* (Sn), *O. latipes* (OI), *T. nigroviridis* (Tn) e *D. labrax* (DI) sono state incluse nell'allineamento. Ciascun gene è definito dall'abbreviazione della specie a cui appartiene seguita dal numero attribuitogli nella specie stessa. A ciascun nodo è stato assegnato il rispettivo valore di *bootstrap*; i valori al di sotto di 50 non sono stati riportati. A lato sono indicate le espansioni "Intronless" specie-specifiche.

Siti aminoacidici dei NITR di medaka (espansione “*Intronless*”) identificati essere sotto selezione positive.

Protein domain ^b	Maximum Likelihood Method			
	PAML ^c	FEL ^d	REL ^e	MEC
V		0.13	37.1	
V			30.6	
V		0.07		
V_CDR1		0.00	5150.6	
V_CDR1			24.1	
V_CDR1	0.990		22.9	[2.5, 2.5]
V_CDR2	0.999			[2.5, 2.5]
V_CDR2			21.3	
V		0.10		
V		0.12		
V		0.19		
V_CDR3	0.993		215.0	[1.1, 2.5]
V_CDR3	0.998			[2.5, 2.5]
V		0.10		
V/C2		0.14		
V/C2		0.19		
V/C2		0.07		
V/C2		0.09		
TM		0.04	40.0	
TM		0.10		
TM		0.04	40.4	
Cyt1		0.06		
Cyt2		0.01		
Cyt2		0.13		

^aV: dominio V, V_CDR1: *Complementarity determining region 1*, V_CDR2: *complementarity determining region 2*, V_CDR3: *Complementarity determining region 3*, V/C2: dominio I, TM: regione transmembrana, Cyt1 e Cyt2: domini citoplasmatici.

^b Probabilità associate all'analisi con BEB.

^c *P-value*.

^d *Bayes factor*.

^e Intervallo di confidenza.

Siti aminoacidici dei NITR di medaka (espansione “*Intronless*”) identificati essere sotto selezione positive.

Protein domain	Maximum Likelihood Method			
	PAML	FEL	REL	MEC
Signal P.		0.05	119.9	
Signal P.		0.18	21.2	
Signal P.		0.09	47	
V		0.17	22.5	
V		0.15	57.3	
V		0.17	23.3	
V_CDR1		0.17		
V_CDR1		0.03	62.7	
V_CDR1	0.999			[1.6, 4.4]
V_CDR1	0.960	0.10		
V_CDR1			23.7	
V		0.12	34.2	
V	0.998			
V_CDR2		0.10	24.8	
V_CDR2	0.998			
V_CDR2	1.00	0.01		[4.4, 4.4]
V_CDR2	0.981	0.11		
V_CDR2	0.997			[1.6, 4.4]
V_CDR2	0.969			
V_CDR2	0.989			
V	0.998	0.09		[1.6, 4.4]
V	0.997			[1.6, 4.4]
V			20.5	
V		0.14		
V_CDR3	0.981			
V	0.971	0.09		
V/C2		0.04	129.9	
V/C2		0.19		
V/C2		0.19		
V/C2	0.999	0.19		[4.4, 4.4]
V/C2		0.03	638.2	
TM	0.981			
TM		0.14		
TM		0.13		
TM		0.03	217.1	
TM		0.06	38.3	
TM		0.07	89.9	
TM	0.993			
TM	0.977			
TM			33.5	
TM		0.11	47.9	
TM	0.982			
Cyt2		0.04	82.5	

APPENDICE 2E

Espressione dei geni NITR valutata mediante RT-PCR in Quattro esemplari di branzino. Per ciascun tessuto, i dati di espressione sono riportati in termini di presenza (+) o assenza (-).

L'espressione dei geni NITR29 e NITR30 è stata osservata solo nel rene cefalico.

NITR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19	20	21	22	23	24	25	26	28	
Rene cefalico	+++	+---	+++	+++	+++	---	---	+++	+++	+++	+++	+++	---	+++	+++	---	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
Sangue	---	---	---	---	---	---/	---	---	---/	---	---	---/	---	++/	+++	---	---	---	---	+	---	---	---	---	---	---	---
Branchie	---	---	+++	+++	+++	+++	---	+++	+++	+++	+++	+++	---	+++	+++	---	+++	---	+++	---	+++	+++	+++	+++	+++	+++	---
Fegato	---	---	---	---	---	---	---	---	---	---	---	---	---	---	+++	---	---	---	---	---	+++	+++	---	---	---	---	---
Intestino	---	---	+++	---	---	+++	---	+++	+++	+++	+++	+++	---	---	+++	---	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
Milza	---	---	+++	---	+++	+++	---	+++	---	+++	+++	+++	---	+++	+++	---	---	+++	---	+++	+++	+++	+++	+++	+++	+++	+++
Pelle	---	---	---	---	---	---	---	---	---	+++	---	+++	---	+++	---	---	+++	---	---	---	+++	---	+++	+++	+++	---	+++
Muscolo	---	---	---	---	---	---	---	---	---	---	---	---	---	+++	---	---	---	---	---	---	---	---	---	---	---	---	---

/: non determinato

APPENDICE 2E

Correlazione di Spearman dei livelli di espressione dei geni NITR di branzino per tutti i possibili confronti a coppie

	NITR10	NITR11	NITR12	NITR13	NITR14	NITR15	NITR1	NITR2	NITR3	NITR4	NITR5	NITR6	NITR8	NITR9	NITR16	NITR17	NITR19	NITR20	NITR21	NITR22	NITR23	NITR24	NITR25	NITR26	NITR28	
NITR10	1.00																									
NITR11	0.74**	1.00																								
NITR12	0.67**	0.54**	1.00																							
NITR13	0.15	0.28	0.24	1.00																						
NITR14	0.28	0.28	0.33	0.10	1.00																					
NITR15	-0.15	0.22	-0.14	0.12	-0.14	1.00																				
NITR1	0.43**	0.55**	0.38*	0.54**	-0.15	0.15	1.00																			
NITR2	0.15	0.15	0.14	-0.12	0.14	0.03	0.22	1.00																		
NITR3	0.75**	0.75**	0.70**	0.36*	0.18	0.19	0.52**	0.17	1.00																	
NITR4	0.21	0.21	0.20	-0.17	0.20	0.05	0.05	0.70**	0.24	1.00																
NITR5	0.60**	0.47**	0.55**	0.08	0.42*	0.16	0.17	0.20	0.45**	0.29	1.00															
NITR6	0.64**	0.51**	0.47**	0.17	0.07	0.14	0.15	-0.14	0.60**	0.07	0.62**	1.00														
NITR8	0.51**	0.51**	0.47**	0.03	0.47**	0.14	0.15	0.23	0.60**	0.33	0.36*	0.33	1.00													
NITR9	0.56**	0.42*	0.52**	0.13	0.10	0.12	0.27	0.27	0.63**	0.38*	0.49**	0.45**	0.45**	1.00												
NITR16	0.21	0.21	-0.07	-0.17	-0.33	0.05	0.31	-0.05	0.24	-0.07	-0.23	0.07	0.07	-0.17	1.00											
NITR17	0.68**	0.43*	0.51**	-0.01	0.12	-0.22	0.22	0.22	0.65**	0.31	0.17	0.28	0.54**	0.54**	0.31	1.00										
NITR19	0.63**	0.37*	0.44*	0.09	0.18	-0.17	0.27	0.17	0.50**	-0.02	0.45**	0.34	0.21	0.36*	0.24	0.39*	1.00									
NITR20	0.68**	0.55**	0.51**	0.27	0.25	0.15	0.22	0.22	0.65**	0.31	0.68**	0.67**	0.41*	0.82**	-0.21	0.35*	0.39*	1.00								
NITR21	0.19	0.19	0.23	0.12	0.23	-0.11	0.23	0.11	0.25	-0.13	0.13	-0.09	0.05	0.12	-0.13	0.09	0.25	0.09	1.00							
NITR22	0.40**	0.40*	0.31	0.16	0.03	0.27	0.28	0.12	0.45*	0.17	0.46**	0.24	0.10	0.45**	-0.10	0.28	0.31	0.42*	0.48**	1.00						
NITR23	0.35*	0.22	0.41*	0.28	-0.12	-0.15	0.43*	0.15	0.37*	0.21	0.22	0.12	0.12	0.28	-0.05	0.30	0.24	0.17	0.19	0.40*	1.00					
NITR24	0.87**	0.61**	0.53**	0.20	0.39*	-0.13	0.33	0.13	0.64**	0.19	0.51**	0.56**	0.42*	0.49**	0.19	0.60**	0.64**	0.60**	0.13	0.22	0.21	1.00				
NITR25	0.81**	0.67**	0.47**	0.10	0.20	-0.14	0.25	0.14	0.70**	0.20	0.42*	0.60**	0.47**	0.52**	0.20	0.64**	0.44*	0.64**	0.23	0.45**	0.15	0.66**	1.00			
NITR26	0.61**	0.47**	0.25	0.35	0.12	0.25	0.33	0.13	0.51**	0.19	0.51**	0.42*	0.29	0.49**	-0.08	0.33	0.51**	0.60**	0.13	0.51**	0.34	0.58**	0.39*	1.00		
NITR28	0.87**	0.74**	0.54**	0.15	0.28	-0.15	0.43*	0.15	0.63**	-0.05	0.47**	0.51**	0.38*	0.42*	0.21	0.55**	0.63**	0.55**	0.33	0.40*	0.22	0.74**	0.81**	0.47**	1.00	

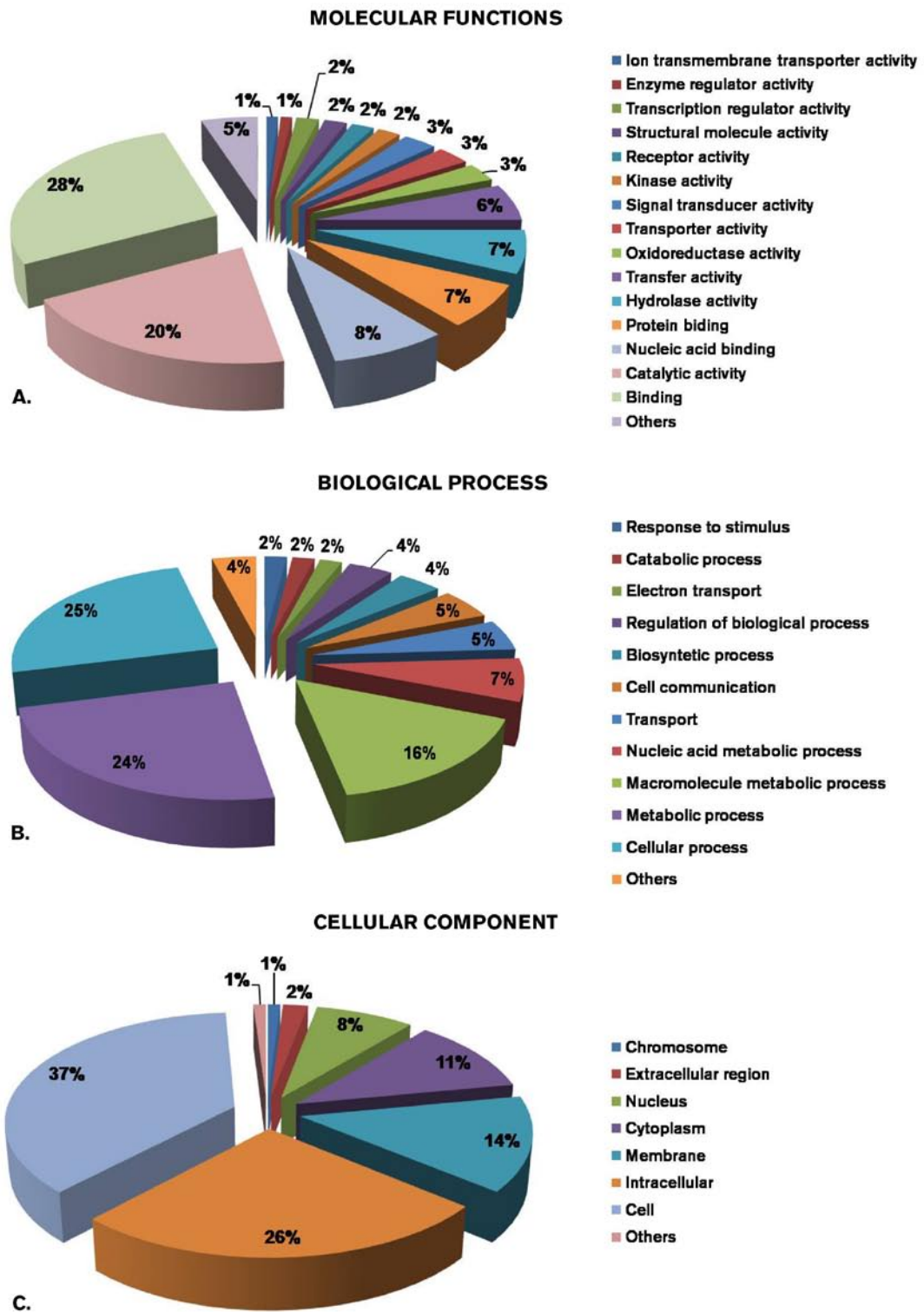
** p<0.01

* p< 0.05

APPENDICE 3A

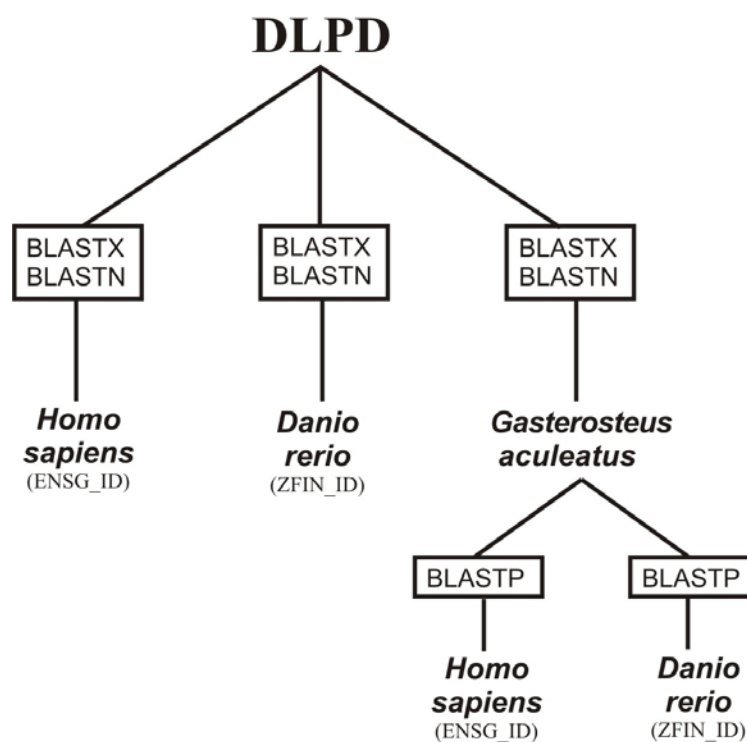
Coppie di *primer* gene-specifiche e sonda UPL corrispondente, disegnate per gli 11 geni *target* selezionati per la validazione in *Real-time* RT-PCR dei dati di *microarray*

<i>SAPD ID</i>	<i>Primer Forward</i>	<i>Primer Reverse</i>	<i>Dimens ampl.</i>	<i>Sonda UPL</i>
<i>Target</i>				
SAPD00597	TGGATCGTCGGAGAGCAC	AACTCCAGCAACATTGACACC	66	47
SAPD26496	GTTGGACACGTCCAATAAGA	GTACTTGCTCTCGCCAGAC	120	20
SAPD02277	TTATCAAAGTGGACCCGTCTG	GGAAGGCACAAGCAAGGTAG	66	8
SAPD04884	GATCTTGATGCTGTGTGTTGG	TGCTGCTGGACTGAGTTCATA	115	51
SAPD10294	GGACATGACCAAGGTGGAGT	GGCAGACGGCTTCTCAA	127	10
SAPD03464	AAGGTCAGCCATGTGAGCA	AACACCGGGCAGGACTTTA	72	67
SAPD19150	CCGTGATTAAGTTCGCAGTG	GATGGCAGACGTCCAACAG	105	14
SAPD01126	TACGACCTGCACCCATTA	GCGACAGCGTAGTAACAGGTC	68	73
SAPD19202	TTGGCTTCACTATGTACACCACTT	TGGTCAGGGTTGTTGCTGA	111	20
SAPD02358	AGCTCCAGATCAGCAGAGA	CTTGGTCTGCACGTCGTCT	87	56
SAPD01680	GCCACACCGATTGTCAGAT	TGAAGAGCCACCACTGTGTC	86	74
<i>Reference</i>				
SAPD02236	CCGTCTGGACCACAACAGA	GCCCCAGATGATCACATTCT	88	87



Termini GO, associate ai trascritti unici di orate, maggiormente rappresentati all'interno delle tre categorie: **A.** "Molecular function", **B.** "Biological process", **C.** "Cellular component".

Schema raffigurante i 4 diversi approcci utilizzati per recuperare gli “identificatori”, utilizzati da David (*Ensembl Human Gene ID* o ZFIN ID), corrispondenti a ciascun trascritto DLPD



Numero di trascritti DLPD per i quali sono stati recuperati gli “identificatori” per David (*Ensembl Human Gene ID* o ZFIN ID) utilizzando i quattro diversi approcci raffigurati sopra.

	ENSEMBL HUMAN GENE ID		ZFIN ID	
	DLPD-GA-HUMAN	DLPD-HUMAN	DLPD-GA-DR	DLPD-DR
N° di DLPD con identificatori in DAVID (ID unici)	9,156 (6,019)	9,277 (6,458)	7,217 (4,955)	7,010 (5,148)
N° di identificatori caricati come “background”	5,385	5,654	4,058	4,150
N° di DLPD significativi con identificatori in DAVID (ID unici)	78 (75)	70 (68)	65 (63)	56 (54)
N° di identificatori caricati come “gene list”	71	60	55	46

APPENDICE 4B

Coppie di *primer* gene-specifiche disegnate per i 13 geni *target* selezionati per la validazione in *Real-time* RT-PCR dei dati di *microarray*

DLPD	PRIMER FORWARD	PRIMER REVERSE
DLPD11717	CCCAGCGGAGGATGTAATAA	GTCCTTCTGCTCCGAGTCTG
DLPD11576	TTTTTGTGAGCCGTCAATCA	CTCCTGGTGGACACTGTGG
DLPD03840	CCACACCTCTTCTTGGCACT	AGCTGCACTAAACACCTGCAT
DLPD07155	CCTCTTCGGCAAGGACATTA	CGCTGTTTTGTGCTTCTCA
DLPD09843	GTGGTGCTGGTACAGTGCAG	AGTCTGTGCAGAGGGACAGG
DLPD13372	GAGTCCAGGAGGCAGATTCA	CTGCCTCAGCAGAGCTTCAT
DLPD08826	CTGTGTGAGCACCAGAGCA	CGATGATGAAGTTTTGTCACTCA
DLPD03006	TCCTCTATCCGCATCTTGCT	GGCAGGGCATCAGTAGAGAG
DLPD03024	TCACAGGGCTCCTCTGGTAG	GGCCTGATCTCCTTTCCAA
DLPD06176	TACATGAAAGCACTCGGAGTG	CCCTCCTGGCTGATGATTA
DLPD02075	CGCACTCTGAAAACATTCCA	AGGGCTCAGTTGTGGAAAAA
DLPD12789	GACTGCAAGCAGACCATCAA	TACTTGCATTCCCCTCCAAA
DLPD06508	TCATGGAGAAGGCTGATTCC	GTTTAGCCACTGCCCAACAT

APPENDICE 4C

Termini GO, associati ai trascritti unici di branzino, maggiormente rappresentati all'interno delle tre categorie: **A.** “*Molecular function*”, **B.** “*Biological process*”, **C.** “*Cellular component*”.

A. MOLECULAR FUNCTION

GO:0003674 Molecular function			
GO Class ID	Definitions	Counts	Fraction
GO:0005488	binding	3795	28.52%
GO:0003824	catalytic activity	2156	16.20%
GO:0005515	protein binding	1160	8.72%
GO:0016787	hydrolase activity	781	5.87%
GO:0003676	nucleic acid binding	665	5.00%
GO:0000166	nucleotide binding	657	4.94%
GO:0016740	transferase activity	643	4.83%
GO:0003677	DNA binding	365	2.74%
GO:0005215	transporter activity	304	2.28%
GO:0016301	kinase activity	282	2.12%
GO:0004871	signal transducer activity	254	1.91%
GO:0008233	peptidase activity	222	1.67%
GO:0030528	transcription regulator activity	198	1.49%
GO:0004872	receptor activity	192	1.44%
GO:0005198	structural molecule activity	185	1.39%
GO:0004672	protein kinase activity	180	1.35%
GO:0003723	RNA binding	145	1.09%
GO:0008289	lipid binding	142	1.07%
GO:0005509	calcium ion binding	123	0.92%
GO:0030234	enzyme regulator activity	116	0.87%
GO:0005102	receptor binding	113	0.85%
GO:0003700	transcription factor activity	97	0.73%
GO:0008092	cytoskeletal protein binding	94	0.71%
GO:0045182	translation regulator activity	63	0.47%
GO:0008135	translation factor activity, nucleic acid binding	61	0.46%
GO:0004721	phosphoprotein phosphatase activity	59	0.44%
GO:0003779	actin binding	57	0.43%
GO:0005216	ion channel activity	52	0.39%
GO:0004518	nuclease activity	38	0.29%
GO:0003774	motor activity	29	0.22%
GO:0030246	carbohydrate binding	26	0.20%
GO:0003682	chromatin binding	21	0.16%
GO:0016209	antioxidant activity	17	0.13%
GO:0019825	oxygen binding	7	0.05%
GO:0005326	neurotransmitter transporter activity	6	0.05%

B. BIOLOGICAL PROCESS

GO:0008150 Biological Process			
GO class ID	Definition	Count	Fraction
GO:0008152	metabolism	2749	21.82%
GO:0019538	protein metabolism	992	7.87%
GO:0009058	biosynthesis	857	6.80%
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolism	800	6.35%
GO:0006810	transport	738	5.86%
GO:0007275	development	689	5.47%
GO:0007154	cell communication	536	4.25%
GO:0007165	signal transduction	468	3.71%
GO:0016043	cell organization and biogenesis	433	3.44%
GO:0006464	protein modification	375	2.98%
GO:0006350	transcription	368	2.92%
GO:0009056	catabolism	307	2.44%
GO:0009653	morphogenesis	304	2.41%
GO:0030154	cell differentiation	215	1.71%
GO:0006996	organelle organization and biogenesis	214	1.70%
GO:0006950	response to stress	213	1.69%
GO:0006412	protein biosynthesis	210	1.67%
GO:0015031	protein transport	193	1.53%
GO:0005975	carbohydrate metabolism	192	1.52%
GO:0006629	lipid metabolism	173	1.37%
GO:0009790	embryonic development	155	1.23%
GO:0006811	ion transport	154	1.22%
GO:0006519	amino acid and derivative metabolism	129	1.02%
GO:0007049	cell cycle	127	1.01%
GO:0006091	generation of precursor metabolites and energy	124	0.98%
GO:0016265	death	110	0.87%
GO:0008219	cell death	110	0.87%
GO:0006259	DNA metabolism	106	0.84%
GO:0007010	cytoskeleton organization and biogenesis	76	0.60%
GO:0019725	cell homeostasis	67	0.53%
GO:0009605	response to external stimulus	65	0.52%
GO:0008283	cell proliferation	60	0.48%
GO:0007267	cell-cell signaling	40	0.32%
GO:0000003	reproduction	39	0.31%
GO:0040007	growth	39	0.31%
GO:0009607	response to biotic stimulus	34	0.27%
GO:0009628	response to abiotic stimulus	32	0.25%
GO:0007610	behavior	27	0.21%
GO:0009719	response to endogenous stimulus	23	0.18%
GO:0016049	cell growth	17	0.13%
GO:0007005	mitochondrion organization and biogenesis	13	0.10%
GO:0040029	regulation of gene expression, epigenetic	11	0.09%
GO:0016032	viral life cycle	8	0.06%
GO:0019748	secondary metabolism	5	0.04%
GO:0008037	cell recognition	3	0.02%
GO:0007028	cytoplasm organization and biogenesis	1	0.01%

C. CELLULAR COMPONENT

GO:0005575 Cellular component			
GO Class ID	Definitions	Counts	Fraction
GO:0005623	cell	3949	36.44%
GO:0005622	intracellular	2729	25.18%
GO:0005737	cytoplasm	1500	13.84%
GO:0005634	nucleus	658	6.07%
GO:0005886	plasma membrane	271	2.50%
GO:0005739	mitochondrion	266	2.45%
GO:0005856	cytoskeleton	230	2.12%
GO:0005576	extracellular region	214	1.97%
GO:0005783	endoplasmic reticulum	161	1.49%
GO:0005829	cytosol	150	1.38%
GO:0005840	ribosome	149	1.37%
GO:0005794	Golgi apparatus	137	1.26%
GO:0005694	chromosome	67	0.62%
GO:0005654	nucleoplasm	60	0.55%
GO:0005768	endosome	44	0.41%
GO:0016023	cytoplasmic membrane-bound vesicle	43	0.40%
GO:0005615	extracellular space	37	0.34%
GO:0005730	nucleolus	28	0.26%
GO:0005773	vacuole	26	0.24%
GO:0005578	extracellular matrix (sensu Metazoa)	24	0.22%
GO:0005764	lysosome	23	0.21%
GO:0005815	microtubule organizing center	22	0.20%
GO:0005635	nuclear membrane	17	0.16%
GO:0005777	peroxisome	12	0.11%
GO:0000228	nuclear chromosome	12	0.11%
GO:0005929	cilium	5	0.05%
GO:0005811	lipid particle	3	0.03%

