



UNIVERSITÀ POLITECNICA DELLE MARCHE
DIPARTIMENTO DI SCIENZE DELLA VITA E DELL'AMBIENTE

Corso di Laurea Magistrale in Biologia Molecolare e Applicata
Curriculum: Tecnologie Biologiche

Un approccio molecolare e bioinformatico per caratterizzare
il recettore della vitellogenina nell'anfibio *Cynops orientalis*

A combined molecular and bioinformatic approach to characterize the
vitellogenin receptor in the amphibian *Cynops orientalis*

Candidata:
Travaglini Elena

Relatore:
Chiar.mo Prof. Barucca Marco

Correlatrice:
Dott.ssa Carducci Federica

Sessione Straordinaria
Anno Accademico: 2020/2021

Con Amore

ai miei genitori, la mia forza.

1. Scopo della tesi	4
2. Introduzione	6
2.1 <i>Gli anfibi e la Cynops orientalis</i>	7
2.2 <i>Biologia della Cynops orientalis</i>	11
2.3 <i>La vitellogenina</i>	15
2.4 <i>I recettori della vitellogenina</i>	25
3. Materiali e Metodi	31
3.1 <i>Campionamento</i>	32
3.2 <i>Estrazione dell'RNA (metodo Trizol)</i>	32
3.3 <i>Quantizzazione dell'RNA</i>	35
3.4 <i>Sequenziamento NGS</i>	37
3.5 <i>Ricerca delle sequenze dei recettori della vitellogenina e confronto nelle banche dati</i>	38
3.6 <i>Completamento delle sequenze dei recettori della vitellogenina</i> ...40	
3.6.1 <i>Retrotrascrizione</i>	40
3.6.2 <i>Progettazione dei primer</i>	43
3.6.3 <i>Rapid Amplification cDNA ends (RACE)</i>	44
3.6.4 <i>Estrazione e purificazione del prodotto di PCR</i>	51
3.6.5 <i>Sequenziamento di Sanger</i>	54

3.7 <i>Analisi dell'espressione genica</i>	55
3.7.1 <i>Mapping</i>	55
3.7.2 <i>Creazione del subset di ortologi per la normalizzazione</i>	55
3.7.3 <i>Calcolo dei valori di espressione</i>	56
3.8 <i>Analisi filogenetica dei dati</i>	57
3.9 <i>Predizione di struttura terziaria</i>	63
4. Risultati	65
4.1 <i>Ottenimento delle sequenze del recettore della vitellogenina dal database</i>	66
4.2 <i>Progettazione dei primer</i>	71
4.3 <i>Completamento della seq1</i>	74
4.4 <i>Analisi dell'espressione genica</i>	81
4.5 <i>Analisi filogenetica</i>	84
4.6 <i>Predizione di struttura terziaria</i>	87
5. Discussione	93
6. Ringraziamenti	102
7. Bibliografia	104

1. Scopo della tesi

1. Scopo della tesi

Nei tetrapodi le informazioni riguardanti i geni della vitellogenina e i loro recettori sono limitati a pochi taxa. Espandere queste conoscenze permette non solo di comprendere a fondo il meccanismo della vitellogenesi ma anche di individuare nuove funzioni e processi metabolici in cui questa proteina è coinvolta. Per fare ciò è indispensabile individuare i suoi recettori, i loro livelli di espressione ed investigare la loro struttura. Ad oggi, per quanto riguarda gli anfibi, studi dettagliati sono stati condotti solo su organismi appartenenti all'ordine Anura. In questa tesi ci si è posti l'obiettivo di identificare il recettore della vitellogenina all'interno del trascrittoma dell'anfibio *Cynops orientalis*, appartenente all'ordine Caudata. La *C. orientalis* è un organismo caratterizzato da un'elevata dimensione genomica, la cui stima è di 44.27 Gb (circa 13 volte il genoma umano). Questo rappresenta un ostacolo significativo per l'ottenimento di un assemblaggio genomico completo. Il trascrittoma è una risorsa molecolare che costituisce una valida alternativa utile per studi di espressione genica, trascrittomica comparativa e analisi evolutive su larga scala. Ciò ha permesso di identificare non solo il recettore della vitellogenina in questo anfibio, ma anche di avere un'identificazione iniziale di altri recettori appartenenti alla famiglia dei recettori per le lipoproteine a bassa densità.

2. Introduzione

2.1 Gli anfibi e la *Cynops orientalis*

Gli anfibi sono vertebrati appartenenti al phylum dei Chordata e sono divisi in tre ordini: Anura, comprendenti rane e rospi, Gymnophiona (o Apoda) di cui fanno parte le cecilie e Caudata (o Urodela) che raggruppa salamandre e tritoni (fig.1).

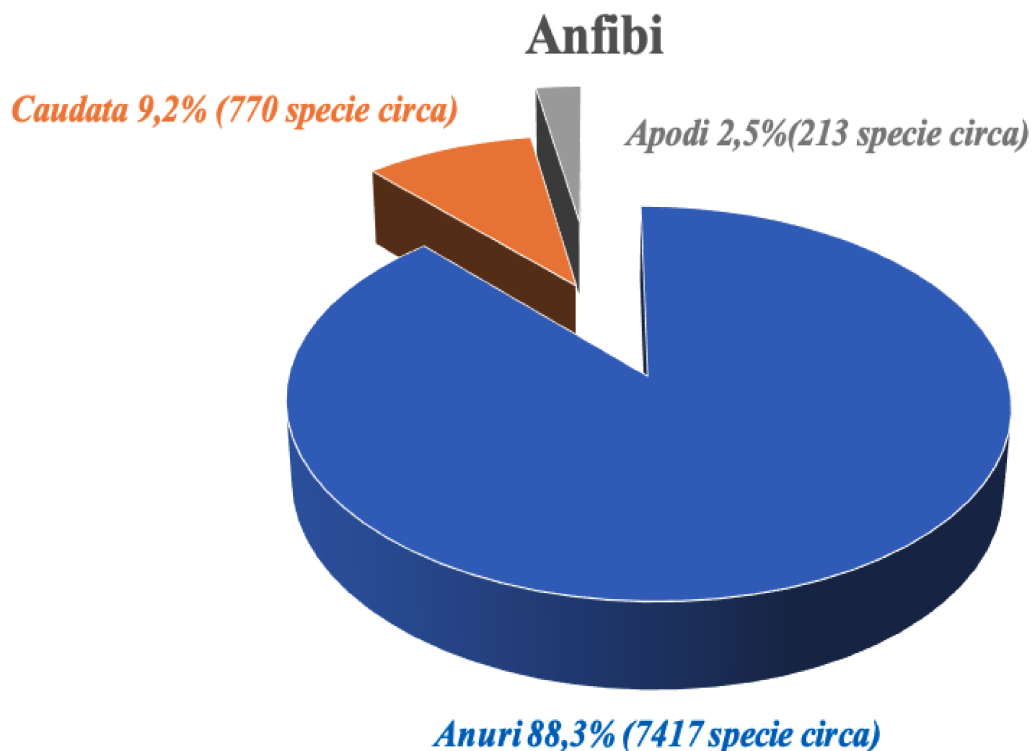


Figura 1. Grafico del numero di specie dei tre ordini di anfibi conosciute ad oggi, registrate dal "American Museum of Natural History" al link <https://amphibiansoftheworld.amnh.org>

La maggior parte di questi organismi sono di piccole dimensioni, ectodermi, capaci di avere un'efficienza di produzione secondaria molto alta: le energie immagazzinate non vengono utilizzate per il controllo termico ma impiegate esclusivamente per la produzione e il mantenimento della propria massa

(Harvey et al., 2007). La possibilità di combinare quindi la piccola dimensione ed un'alta efficienza nella produzione della propria biomassa, ha permesso agli anfibi di occupare un ruolo unico nell'ecosistema, consumando prede troppo piccole per essere mangiate da altri vertebrati (Pough et al., 1980).

Hanno una pelle priva di squame e ricoperta di muco che rappresenta la sede principale di scambio tra ossigeno e anidride carbonica, sia per gli adulti che per le larve (Harvey et al., 2007): proprio per queste caratteristiche, gli anfibi ad oggi disponibili prendono il nome anche di lissanfibi (Lissamphibia, “pelle liscia”). La loro filogenesi è stata per anni controversa e ad oggi la loro origine viene collocata nel Devoniano, il quarto periodo del Paleozoico (circa 400 milioni di anni fa) anche se gli ordini diffusi ai giorni nostri, sopravvissuti all'estinzione di massa alla fine del Cretaceo, risalgono al Mesozoico (Wake et al., 2018). La classe degli anfibi è diffusa praticamente in tutti gli habitat terrestri e strettamente legata a specchi d'acqua dolce, tranne che nelle regioni più fredde e aride o nelle isole oceaniche più remote (Frost et al., 2006). C'è però una differenza nella diffusione dei tre ordini: gli anuri, ad esempio, sono diffusi nella maggior parte delle grandi e piccole isole continentali in regioni temperate e tropicali (Madagascar, Filippine, Borneo e Nuova Guinea) mentre sono assenti nella maggior parte delle isole oceaniche; le salamandre (Caudata) al contrario sono distribuite principalmente nelle regioni settentrionali

(nonostante raggiungano anche il sud America) ma sono assenti nella maggior parte delle grandi isole continentali e nelle isole oceaniche; le cecilie (Apodi) sono invece limitate esclusivamente ai tropici (Wake et al., 2018) (fig.2).

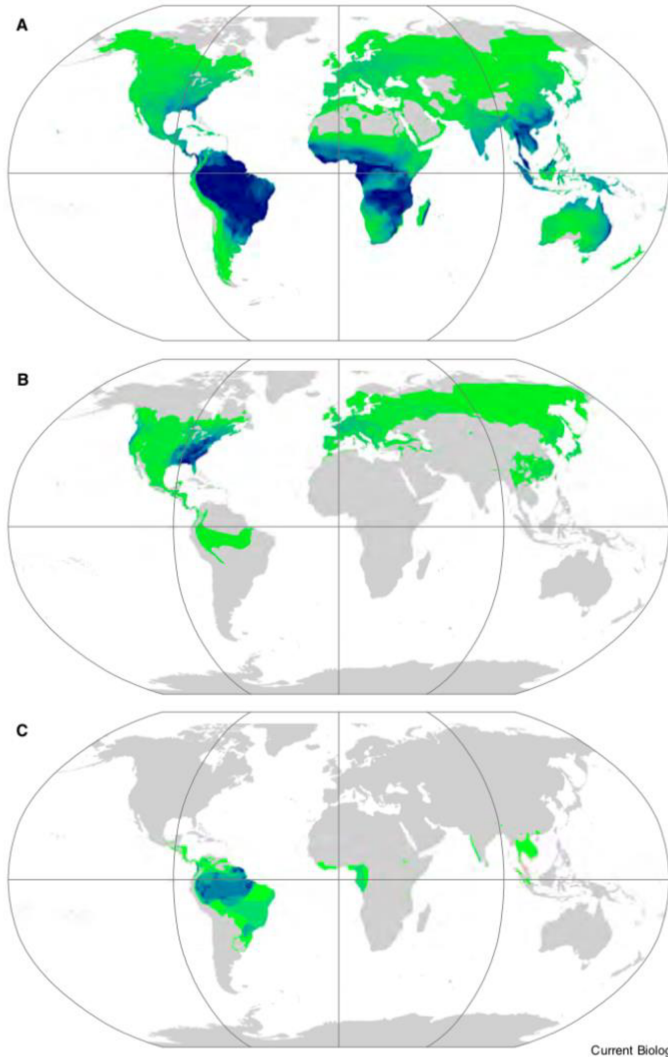


Figura 2. Diffusione degli ordini degli anfibi (Wake et al., 2018): la distribuzione degli Anura (A), dei Caudata (B) e degli Apodi (C). I colori più scuri verso il blu indicano una maggiore ricchezza di specie.

La presenza e distribuzione degli anfibi sta però diminuendo sempre più a causa del cambiamento climatico e per le caratteristiche intrinseche di questi individui stenotermici (Nowakowski et al., 2017): dal 1993 si sono registrate circa 500 popolazioni di rane e salamandre in declino o che necessitano di particolari azioni di conservazione (Blaunstein et al., 2002).

L'ordine dei Caudata è suddiviso in dieci famiglie delle quali quella dei Plethodontidae è caratterizzata dalla maggiore diversità in termini di numero di specie. Quest'ordine è caratterizzato da una superfamiglia monofiletica che è quella delle Salamandroidea. La loro morfologia sembra non aver subito una particolare differenziazione nel corso dell'evoluzione: sembrano essere rimasti invariati a partire dal Giurassico (Gao et al., 2001) e ciò che li distingue dagli altri anfibi è la presenza di una coda anche allo stato adulto e due paia di arti uguali in dimensione posti perpendicolarmente al corpo; non hanno denti su entrambi le mascelle e hanno costole sulla maggior parte delle vertebre (Vieites et al., 2009). Le salamandre sono i tetrapodi con il genoma più grande e, fatta eccezione per i pesci polmonati, anche tra i vertebrati (Sun et al., 2012).

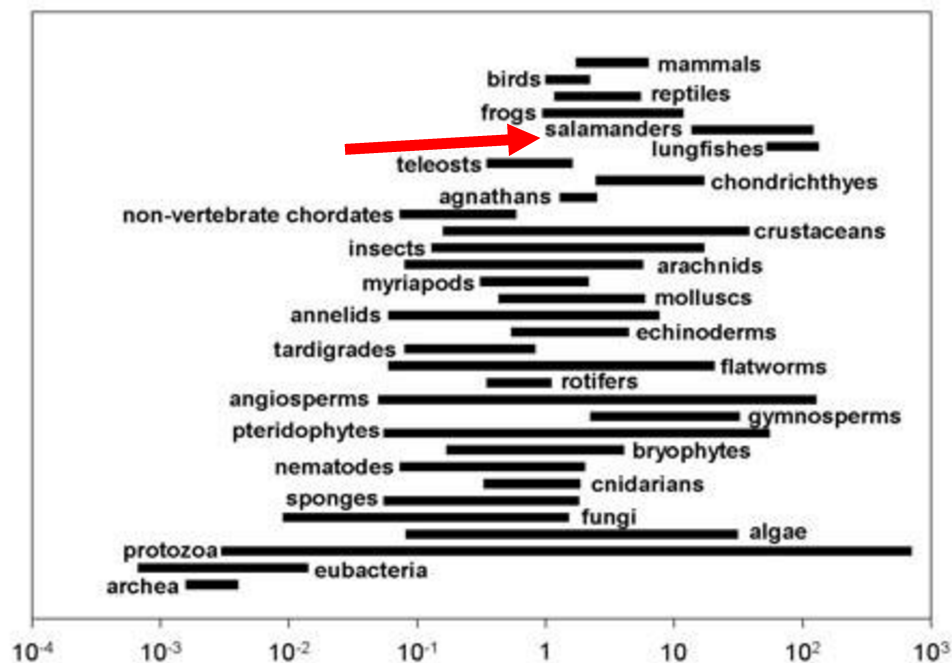


Figura 3. Rappresentazione della grandezza del genoma dei principali organismi invertebrati e vertebrati, espresso in C-value (pg). (www.genomesize.com)

Il loro genoma ha infatti una dimensione che va dai 14 ai 120Gb (www.genomesize.com) ed è rappresentato principalmente da DNA ripetitivo, che tende ad espandersi per l'attività di elementi trasponibili (transposable elements, TEs) (Metcalf et al., 2013).

L'organismo preso in considerazione in questa tesi è la *Cynops orientalis*, la cui classificazione tassonomica è la seguente:

- Ordine: Caudata
- Famiglia: Salamandridae
- Sottofamiglia: Pleurodelinae
- Genere: *Cynops*
- Specie: *Cynops orientalis*

Facendo parte della famiglia delle Salamandridae, anche questo organismo è caratterizzato da un gigantismo genomico (44,22Gb) con un *C-value* di 44,27pg (www.genomesize.com).

2.2 Biologia della Cynops orientalis

La *C. orientalis* è una salamandra endemica della Cina: è stata infatti identificata nelle province di Anhui, Hubei, Hunan, Zhejiang, Jiangxi, Jiangsu

e Fujian, al sud della Cina, ad un'altitudine che va dai 30 ai 1000m sul livello del mare (www.iucnredlist.org) (fig.4).

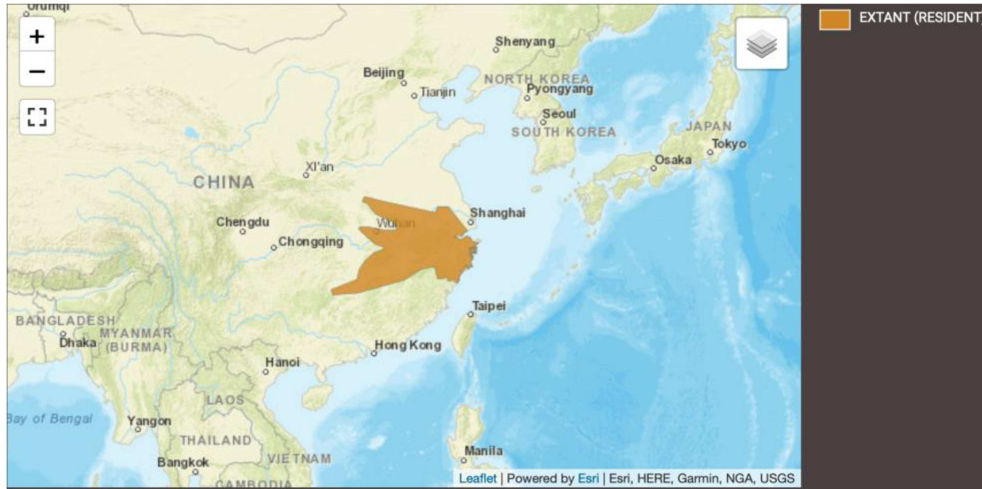


Figura 4. Distribuzione della C. orientalis riportata dalla IUNC RED LIST.

La loro morfologia è molto singolare tanto da renderla facilmente riconoscibile rispetto alle altre specie: sono caratterizzate da una pelle liscia con un dorso nero scuro (fig.5) mentre il lato ventrale ha una colorazione aposematica arancione brillante o rosso (fig.6).



Figura 5. Individuo adulto di C. orientalis fotografato da Jessica Miller 2004.



Figura 6. Colorazione ventrale aposematica della C. orientalis. L'individuo minacciato finge di essere morto per scoraggiare il predatore (tanatosi) (www.amphibiachina.org).

È una specie caratterizzata da un dimorfismo sessuale, con le femmine che possono raggiungere lunghezze di 8,0 cm mentre i maschi raggiungono dimensioni di 6,0-6,5 cm, con una coda più corta ma con una pinna caudale più alta rispetto alle femmine.

Il suo habitat è caratterizzato dalla vicinanza e dalla presenza di ristagni d'acqua dolce, tanto che la si può trovare facilmente nei campi allagati, in piccoli ruscelli o stagni in montagna (Sparreboom et al., 1997), dove la presenza di vegetazione le permette sia di nascondersi sia di deporre le uova: sono individui prevalentemente acquatici ma, occasionalmente, migrano per brevi distanze sul terreno al fine di raggiungere altri ristagni d'acqua. La stagione riproduttiva inizia a marzo e termina a luglio, quando la temperatura dell'acqua è tra i 15 e i 23°C (Lu et al., 2016). I comportamenti durante la loro riproduzione non sono stati osservati in natura, ma sono stati descritti solo in

cattività da Max Sparreboom e colleghi (1997). Le fasi di corteggiamento sono le seguenti:

- *approccio*. Il maschio si avvicina alla femmina e la annusa soprattutto vicino alle ghiandole parotidi.

- *corteggiamento*. Il maschio, durante l'inseguimento, tenta di disporsi davanti alla femmina, perpendicolarmente, flette la coda facendola vibrare, alternando momenti in cui la tiene dritta e immobile, per poi deporre la spermatofora. Può accadere che nella fase di corteggiamento si inserisca anche un altro maschio che può distrarre e allontanare la femmina o stimolare il primo maschio alla deposizione di una spermatofora infruttuosa mimando il comportamento femminile.

- *raccolta della spermatofora*. La femmina poi raccoglie la spermatofora nella sua cloaca. Se l'inseminazione avviene con successo, la femmina deporrà un singolo uovo (al massimo 4) sulla faccia ventrale di una foglia acquatica; può anche piegare la foglia affinché l'uovo risulti più protetto. La femmina in una stagione riproduttiva può arrivare a deporre 100 uova che si schiuderanno nel giro di 13-24 giorni. Le Pleurodelinae hanno uno stato larvale a vita libera (Vitt et al., 2014) e metamorfosando, perdono le branchie esterne, acquistano gli arti, cominciano ad utilizzare i polmoni e raggiungono uno stadio giovanile che sarà sessualmente maturo intorno a 3 anni di vita (www.reptilesmagazine.com).

2.3 La vitellogenina

La vitellogenina (Vtg) è una apolipoproteina multidominio prodotta dalle femmine ovipare e ovovivipare (Verderame et al., 2017; Biscotti et al., 2018) durante la stagione riproduttiva (Verderame et al., 2017). Il nome di questa proteina è dato dalla sua principale funzione che è quella di essere racchiusa all'interno degli oociti prendendo parte alla formazione del vitello, raggiunto tramite i capillari della teca follicolare (Wahli et al., 1981): è quindi una proteina fondamentale nei primi stadi di sviluppo embrionale. Oltre alla funzione di nutrimento durante lo sviluppo embrionale, sono stati annotati differenti ruoli della Vtg: ad esempio, nei teleostei, si è visto che la sua degradazione all'interno degli oociti permette di produrre un pool di amminoacidi che garantisce il gradiente osmotico fondamentale per assicurare l'idratazione dell'oocita, influenzando anche la galleggibilità delle uova pelagiche (Finn et al., 2007; Reading et al., 2009). Poiché è stato osservato che la quantità di Vtg che viene interiorizzata all'interno degli oociti non equivale alla quantità di Vtg disponibili nel circolo sanguigno (Williams et al., 2014), è possibile ipotizzare una funzione differente delle Vtg: queste proteine non sembrano essere solo coinvolte nella formazione del vitello embrionale, bensì hanno un ruolo fondamentale anche nella difesa immunitaria, nella protezione dai radicali liberi fungendo da antiossidante (Sun et al., 2015) (fig.7).

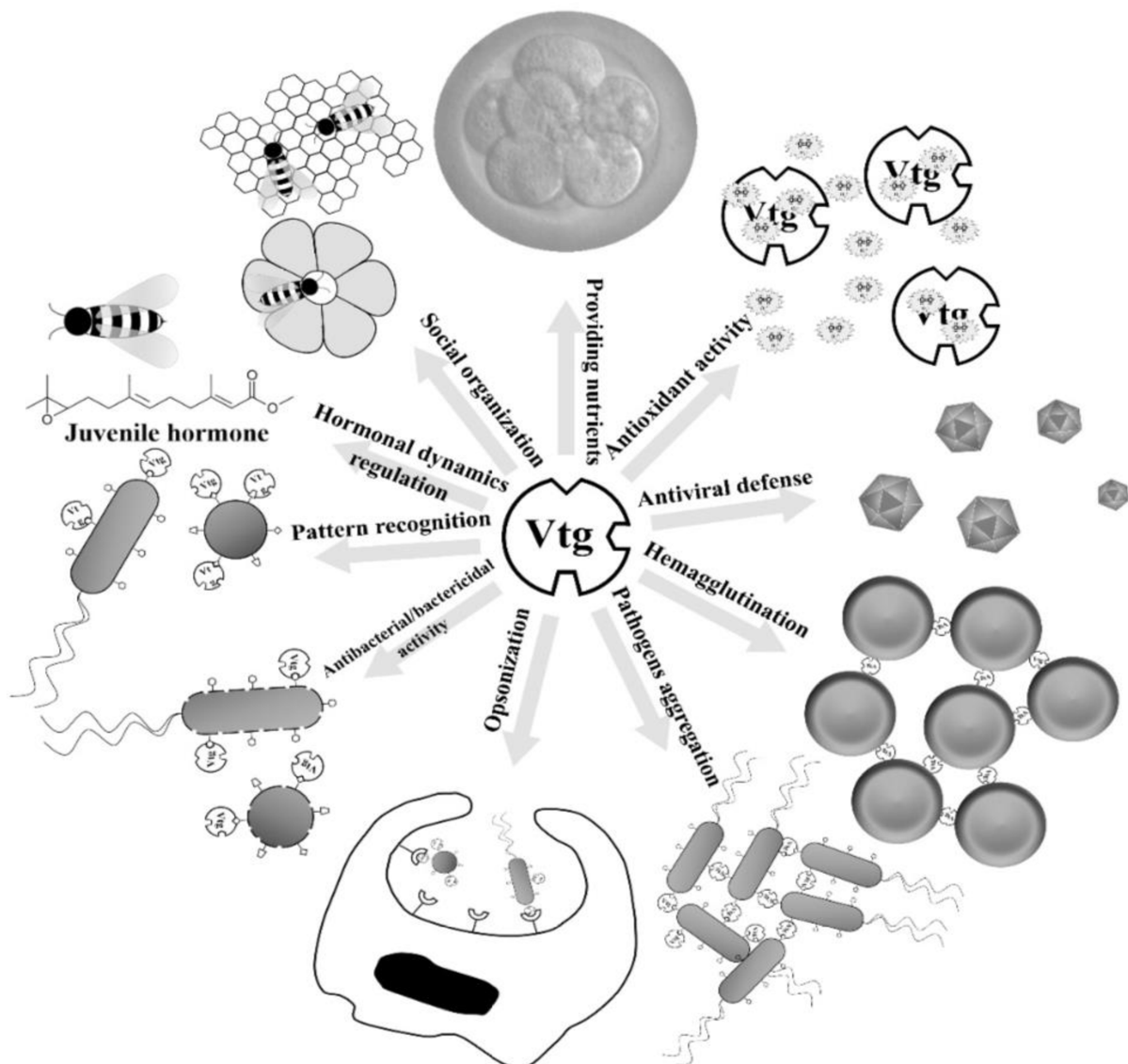


Figura 7. Ruoli multipli di Vtg (Sun et al.,2015): fornisce nutrienti per lo sviluppo di embrioni e larve; influenza l'organizzazione sociale nelle api; riconosce corpi estranei facilitando l'uccisione di batteri o la neutralizzazione dei virus, migliorandone la fagocitosi agendo come opsonina; ha la possibilità di emoagglutinare gli eritrociti e aggregare agenti patogeni e svolge un ruolo antiossidante.

Analisi ELISA (Enzyme-Linked Immunosorbent Assay) hanno dimostrato un'affinità della Vtg con i componenti strutturali identificati come pattern molecolari associati ai patogeni (PAMPs), inclusi i lipopolisaccaridi (LPs) di Gram negativi, l'acido lipoteicoico (LTA) dei Gram positivi, il peptidoglicano

di Gram positivi e negativi e il glucano nei funghi (Li et al., 2008): ciò dimostra come Vtg possa agire come un recettore multivalente, presentando quindi un ruolo anche nella difesa immunitaria dell'ospite fungendo da rivelatore e da opsonina (Sun et al., 2015). Si è anche identificato un sottodominio della Vtg (il dominio D del fattore von Willebrand, *vWFD*) in grado di interagire con la tetraodotossina, una tossina tipica *Takitugu pardalis*: questo legame permette il trasporto della tossina dal fegato all'ovario, dove viene conservata all'interno degli oociti, in modo da fungere da repellente per i predatori ma in grado di attirare i maschi fungendo da ferormone e garantendo la fecondazione delle uova (Yin et al., 2017).

È una proteina che viene espressa principalmente nel fegato femminile, anche se bassi livelli di espressione sono stati riscontrati anche nei maschi (Barucca et al., 2010). In particolare, in quest'ultimi, a seguito di un'esposizione estrogenica o a interferenti endocrini derivati da attività antropiche, si può avere una sovraespressione che in alcune specie come *Podarcis sicula* può anche causare un rallentamento della spermatogenesi, mettendo a rischio la continuità della specie stessa (Verderame et al., 2017). Nelle femmine la produzione è regolata mediante un controllo endocrino coordinato che coinvolge cervello, ovario e fegato: ciò permette una produzione stagionale/ciclica in funzione delle gonadotropine normalmente prodotte (Carducci et al., 2019) garantendo

una preparazione adeguata alla stagione riproduttiva. Sotto lo stimolo della kisspeptina (Smith, 2008) (fig.8) affinché venga secreto l'ormone di rilascio delle gonadotropine (GnRH) nell'area preottica, si ha una produzione degli ormoni follicoli stimolanti (FSH e LH) dall'ipofisi (Kumar et al., 2003): si ha uno stimolo a livello follicolare, della produzione di 17β -estradiolo (E2) che viene riconosciuto da un recettore specifico a livello del fegato.

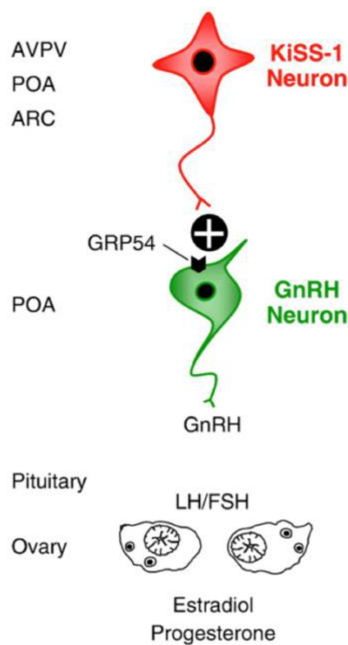


Figura 8. Azione regolatoria proposta della kisspeptina sul GnRH e sull'asse ipotalamo-ipofisi-gonadi proposta da Smith, 2008; nucleo periventricolare anteroventrale (AVPV), area preottica (POA), nucleo arcuato (ARC) e recettore della kisspeptina (GPR54).

L'estrogeno è riconosciuto poi da un recettore nucleare che svolge la sua funzione se forma un dimero con un secondo recettore legato a un secondo estrogeno: solo così riconoscono gli elementi sensibili agli estrogeni (ERE) a monte di sequenze genomiche, tra le quali c'è il gene *vtg*, che viene quindi trascritto e tradotto (Yuji et al., 2020). La Vtg così prodotta subisce delle modifiche post-traduzionali tra cui fosforilazione, glicosilazione e aggiunta di

gruppi lipidici prima di essere rilasciati nel flusso sanguigno come complesso omomero (Carducci et al., 2019). Nello *Xenopus laevis* si è visto che ha un'emivita di circa 40 giorni negli individui maschi e nelle femmine a cui sono stati esportati gli ovari (Wahli et al., 1981). Una volta raggiunto l'ovario, Vtg viene racchiusa nell'ocita e subisce dei tagli proteolitici per la formazione di proteine più corte identificate anche come vitelline o lipovitelline (Lv) e concentrate nel tuorlo (Jorgensen, 2008). La proteina neosintetizzata può essere divisa in diversi componenti (Carducci et al., 2019):

- *LvH*, la catena pesante della Lv che comprende quattro domini: N-sheet, α -elica, C-sheet e A-sheet. È la componente della Vtg che viene riconosciuta dal suo recettore a livello dell'ocita e ne favorisce l'endocitosi (Canapa et al., 2012). In particolare, la sequenza che viene riconosciuta è formata da soli otto amminoacidi a livello del dominio N-sheet (Reading et al., 2009);

- *Pv*, la fosvitina è il secondo dominio della Vtg che sembra essere coinvolta nella sua conservazione all'interno del tuorlo, evitandone la degradazione e ne garantisce anche l'aumento della solubilità durante la crescita dell'ocita avendo la possibilità di interagire con Ca^{2+} , utile per l'osteogenesi (Carducci et al., 2019), P_i e altri metalli (Canapa et al., 2012). Questo dominio possiede un sito di glicosilazione per legare i carboidrati che, insieme agli ioni legati, promuove la solubilità in ambiente acquoso (Carducci et al., 2019). Queste

caratteristiche dipendono esclusivamente dalla sequenza amminoacidica del dominio stesso, che infatti risulta essere ricco di serine, anche fosforilate (Finn et al.,2009)

-LvL, la catena leggera della lipovitellina è un dominio lungo circa 200 amminoacidi, contenente anche un sito di glicosilazione (Carducci et al., 2019);

-vWFD, il dominio D del fattore von Willebrand che contiene una componente β' (β' -c) e una componente terminale (Ct). È il dominio implicato nella dimerizzazione della Vtg.

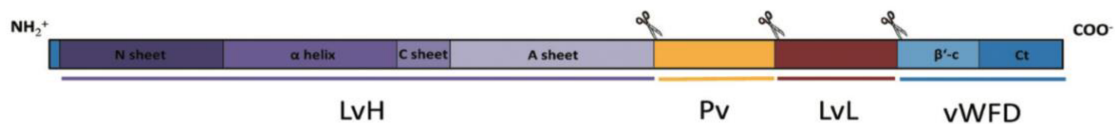


Figura 9. Rappresentazione dei domini della sequenza della vitellogenina presa da Carducci et al., 2019; catena pesante della lipovitellina (LvH), fosvitina (Pv), catena leggera della lipovitellina (LvL) e dominio D del fattore von Willebrand (vWFD). Le forbici rappresentano le posizioni dei siti di taglio riconosciuti.

Tra questi domini sono presenti dei siti di taglio che vengono riconosciuti da specifiche proteine a livello dell'oocita (fig.9): quando Vtg viene endocitata tramite la polimerizzazione di clatrina, si ha un'acidificazione degli endosomi grazie alle pompe protoniche presenti sulla superficie della vescicola, con la conseguente attivazione della catepsina D che, riconoscendo i siti di taglio, taglia Vtg nei costituenti Pv, LvL e vWFD (Hara et al.,2016).

In realtà Vtg non è unica e non è codificata da un singolo gene: si parla di una famiglia genica che nel corso dell'evoluzione e a seconda della storia evolutiva di ogni taxa ha portato a duplicazioni del singolo gene, creando appunto una vera e propria famiglia genica. Il numero delle Vtg presenti per ogni taxa non è conservato poiché nel corso dell'evoluzione ci sono state anche perdite genetiche o duplicazioni che hanno coinvolto i singoli taxa in maniera differente. Inoltre, le varie perdite di materiale genico all'interno di questa famiglia non hanno coinvolto necessariamente l'intera porzione codificante, ma anche solo porzioni di essi, generando quindi dei geni tronchi di alcuni domini, come tutti i geni *vtgC* sequenziate e ottenute dai teleostei (Finn et al., 2007). Data la molteplicità di questa famiglia e la possibilità di avere un'evoluzione differente tra le varie linee evolutive, è molto difficile riuscire a ricostruirne la storia evolutiva (Finn et al., 2009). Nei Sarcopterigi basali, sono state individuate tre geni *vtg* nei celacanti (*Latimeria menadoensis* e *L. chalumnae*) (Canapa et al., 2012) e quattro nei pesci polmonati (*Protopterus annectens*) (Biscotti et al., 2018). Per quanto riguardano i taxa di ordine superiore, i geni per Vtg sono tre negli uccelli e nei rettili, quattro negli anfibi, mentre nei mammiferi ne è stata identificata solo una nell'ornitorinco (*Ornithorhynchus anatinus*) (Biscotti et al., 2018).

Per l'ordine dei Caudata non ci sono particolari informazioni utili riguardo questa famiglia genica: l'unico lavoro disponibile è quello effettuato da Carducci et al. (2021) sulla *C. orientalis* in questo organismo hanno individuato quattro trascritti riconducibili ai geni *vtg*: *vtgCoI*, *vtgCoII*, *vtgCoIII* e *vtgCoIV*. Poiché non è ancora disponibile il suo genoma, le analisi filogenetiche e microsintetiche sono state svolte considerando i trascritti ottenuti da differenti tessuti. Dalle analisi filogenetiche effettuate con le sequenze amminoacidiche dei trascritti, *vtgCoI* risulta essere evolutivamente correlata a quella della cecilia *Microcaecilia unicolor* e la sua presenza in due differenti ordini degli anfibi è dovuto alla sua presenza nell'antenato comune, mentre negli Anuri sembra essere stato persa (fig.10).

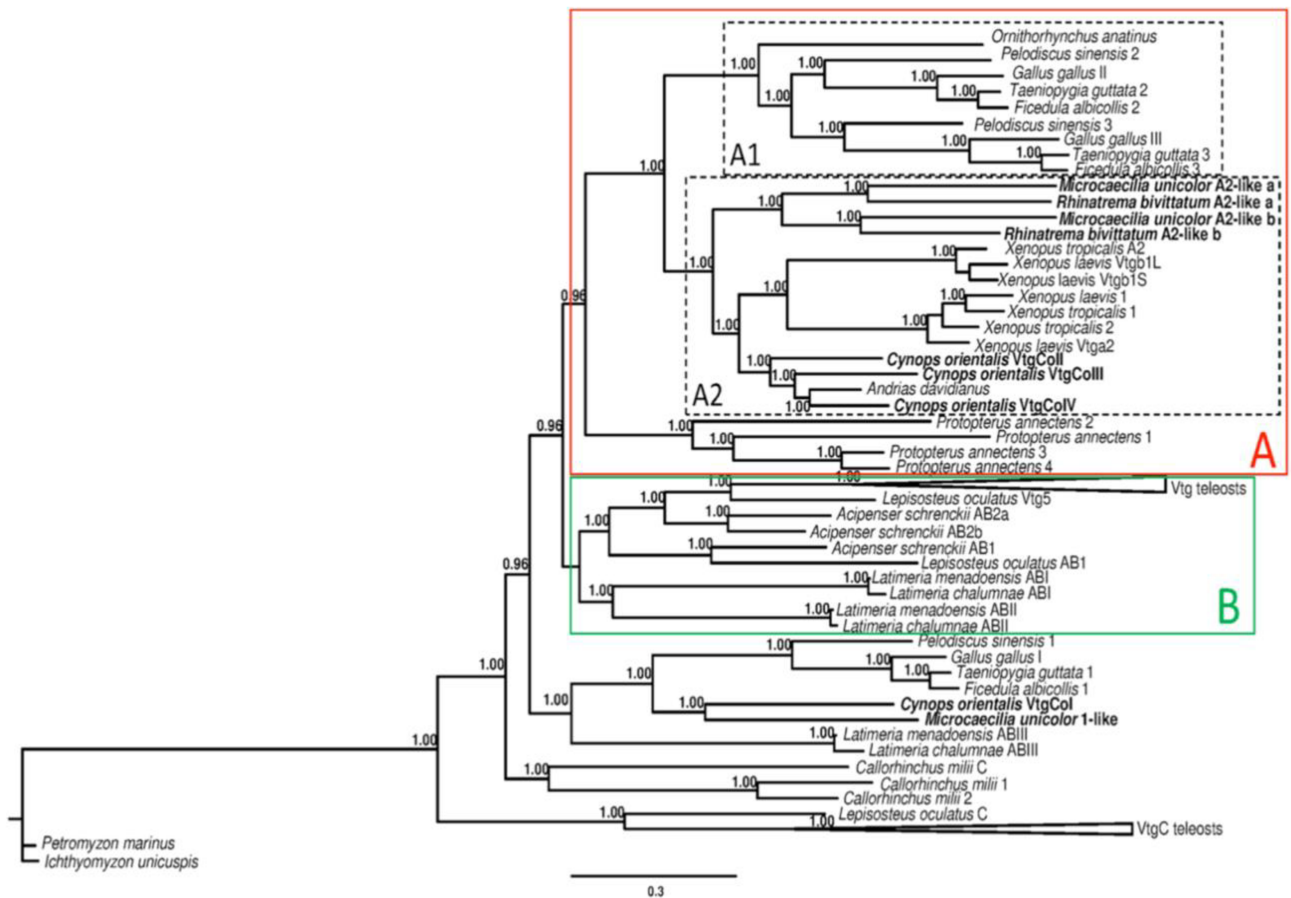


Figura 10. Analisi filogenetica condotta sulle sequenze amminoacidiche delle Vtg identificate in *C. orientalis* pubblicata da Carducci e colleghi (2021).

Per quanto riguarda invece *vtgCoII*, *vtgCoIII* e *vtgCoIV*, si è ipotizzato che si siano formati da duplicazioni indipendenti nella *C. orientalis* a partire da un unico gene ancestrale, dato che dalle analisi filogenetiche le Vtg della salamandra raggruppano insieme piuttosto che con quelle degli altri tetrapodi. In particolare, ci sono due ipotesi che riguardano l'evoluzione di questa famiglia genica all'interno degli anfibi:

-la prima ipotesi prevede un antenato in comune con due *vtg*, *vtgI* e *vtgII/III*, poiché *vtgI* è presente sia nei Gymnophiona che nei Caudata, mentre è stato perso negli Anuri. *vtgII/III* avrebbe poi subito una duplicazione in tutti gli ordini degli anfibi, in modo indipendente, portando ad un numero di tre *vtg* nelle cecilie e quattro nella *C. orientalis*. Negli anuri invece la presenza di due duplicazioni da parte del *vtgII/III* e l'evento di poliploidizzazione avrebbe permesso la generazione di quattro Vtg;

- la seconda ipotesi prevede l'antenato degli anfibi con tre geni: *vtgI*, *vtgII* e *vtgIII*, con il gene *vtgII* e *vtgIII* derivati da una duplicazione ancestrale. Poiché le differenze accumulate da questi due geni non sono particolarmente marcate, nel corso dell'evoluzione sono andati incontro a ricombinazione all'interno dei tre ordini. Quindi negli anfibi, come nei Gymnophiona, ci dovrebbero essere tre geni ma in *C. orientalis* e negli altri urodeli si è verificata poi un'ulteriore duplicazione ulteriore che ha portato alla realizzazione di un quarto gene. Negli anuri invece *vtgI* è andato perso (fig.11).

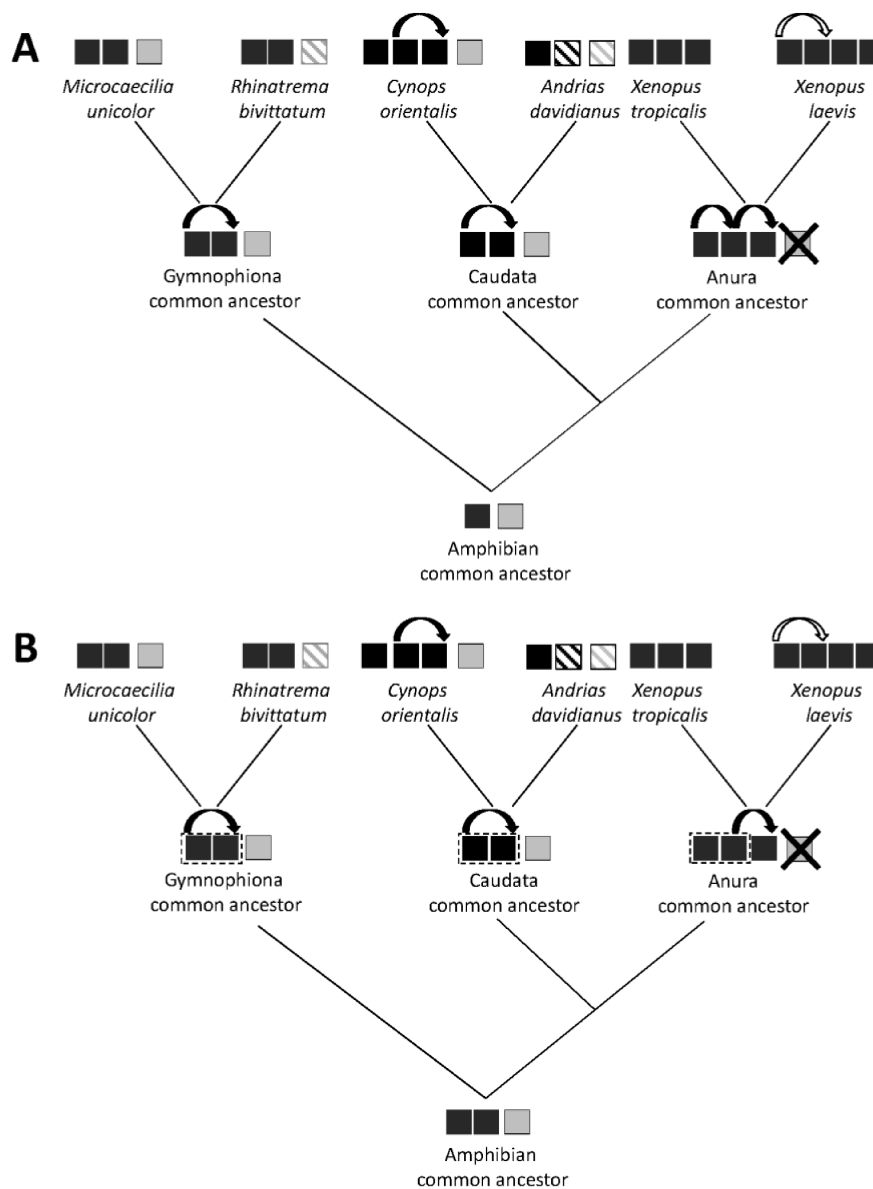


Figura 11. Rappresentazione schematica delle due ipotesi per l'evoluzione delle Vtg negli anfibi di Carducci et al., 2021. Il primo schema (A) mostra l'ipotesi basata sui risultati filogenetici ottenuti; il secondo schema (B) rappresenta la seconda ipotesi che individua anche eventi di conversione genica. I quadrati neri sono i geni Vtg presenti nella regione M; i quadrati grigi rappresentano i Vtg singoli della regione S (VtgI); a strisce sono rappresentati quei geni che possono essere comunque presenti nel genoma, anche se ancora non identificati; la croce su quadrati indica la perdita del gene e le frecce indicano eventi di duplicazione, tranne nello *Xenopus laevis* dove indica un evento di poliploidizzazione.

2.4 I recettori della vitellogenina

Oltre al ruolo della Vtg, è fondamentale caratterizzare e comprendere anche i suoi recettori sia perché sono responsabili della quantità di Vtg che viene incorporata all'interno dell'oocita e che andrà quindi a regolare la crescita e la differenziazione dell'oocita stesso, sia perché il recettore ha un'espressione dipendente dalle caratteristiche ambientali (Wallace et al., 1978). I recettori della vitellogenina (Vitellogenin Receptors, VtgRs) appartengono alla famiglia genica dei recettori per le lipoproteine a bassa densità (Low-density lipoprotein Receptors, LdlRs), (Bujo et al., 1994; Davail et al., 1998) una famiglia di geni che codificano per recettori coinvolti nell'endocitosi di ligandi di differente natura, di cui la maggior parte sono lipoproteine plasmatiche (Reading et al., 2014) coinvolte nel meccanismo lipidico sia di vertebrati che invertebrati (Li et al., 2003). Di questa famiglia fanno parte anche altri recettori come le proteine correlate ai LdlR (Lrp1, Lrp2, Lrp3, Lrp4, Lrp6, Lrp7, Lrp8, Lrp9, Lrp11, Lrp12), il recettore per la sortilina correlato alla Lr11 e le Very low density lipoprotein Receptor (VldlR), dei mammiferi, quest'ultime indicate anche come VtgR nelle altre specie (Reading et al., 2014) o come Lr8 (Morini et al., 2020). Sono recettori che condividono differenti domini, comuni ad ogni tipo di LdlR (Herz, 2001) poiché sono porzioni fondamentali per la loro funzione e poiché derivano da un antenato comune:

- ripetizioni ricche in cisteina all’N-terminale (Goldstein et al., 1985) fondamentali per l’interazione con il ligando (Ligand-Binding Repeats, LBRs), in particolare sono otto nei VldLR/VtgR (Hiramatsu et al., 2015; Mizuta et al., 2013) mentre sono sette per le altre LdlR (Li et al., 2003). Ogni LBR contiene 40 amminoacidi di cui sei sono le cisteine fondamentali per il corretto ripiegamento della porzione, creando tre ponti solfuro intraproteici (Koduri et al., 2001);
- porzioni di cisteine simili al precursore del fattore di crescita dell’epidermide (EGFP-like) distanziate da porzioni povere in cisteina;
- un dominio transmembrana;
- una coda corta al C-terminale rivolta nella porzione citoplasmatica, che contiene una sequenza consenso di serina, aspartato e glutammato (SDE) (Li et al., 2003).

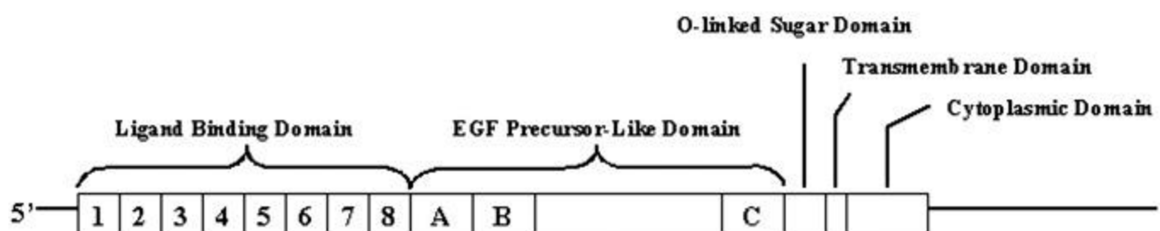


Figura 12. Rappresentazione dei domini del VtgR di Oreochromis aureus, realizzata da Li e colleghi, 2003.

Per caratterizzare le componenti che vengono coinvolte durante il legame con Vtg, Li e colleghi (2003) hanno realizzato dei recettori frammentati nelle

porzioni delle LBR e hanno annotato che solo le proteine che mancavano di una o più porzioni dalla LBR1 alla LBR3 non erano capaci di interagire con Vtg, deducendo quindi come queste componenti siano fondamentali durante il riconoscimento: la porzione della Vtg che viene a legarsi al recettore è stata identificata nella porzione della lipovitellina, al livello del suo N-terminale (Stifani et al., 1990). Il meccanismo principale con cui Vtg viene incorporata negli oociti è stato individuato nell'endocitosi mediata da recettore (Opresko et al., 1987): l'internalizzazione del recettore con il ligando avviene in aree specifiche della membrana dell'oocita, indicate come “*coated pits*” ricoperte da VtgR, che si invaginano verso l'interno della cellula e con la polimerizzazione delle clatrine che permettono il distacco dalla membrana e la realizzazione di un endosoma. Le vescicole si fondono poi formando i corpi multivescicolari all'interno dei quali avviene la scissione della Vtg nelle sue componenti (Patiño et al., 2002). Questi recettori sono stati individuati principalmente sulla superficie degli oociti (Opresko et al., 1987) anche se sono stati identificati bassi livelli di mRNA del VtgR in altri tessuti come nei muscoli e nel cuore del *Gallus gallus* (Bujo et al., 1994).

Anche l'espressione del gene *vtgR* è sotto il controllo ormonale, indotto dall'FSH (Romano et al., 2000) che garantisce la corretta produzione di Vtg e VtgR, favorendo una maturazione e differenziazione dell'oocita ottimale. Nella

trota iridea (*Oncorhynchus mykiss*), l'espressione dell'mRNA del VtgR inizia a livello degli oociti nella fase previtellogenica, diminuendo poi all'inizio della vitellogenesi (Perazzolo et al., 1999): la sua apparente diminuzione nella trascrizione durante l'accumulo del vitello è stata associata a un riciclo del recettore durante le fasi di endocitosi, garantendo quindi un continuo assorbimento senza la necessità di produrre altro recettore (Patiño et al., 2002). In realtà sono stati ipotizzati recettori alternativi (Hiramatsu et al., 2015) in risposta alla vasta famiglia delle Vtg. Oltre ai VtgR/VldLR, infatti, è stato individuato anche un nuovo recettore, l'Lrp13, capace di legare Vtg e garantirne l'endocitosi (Reading et al., 2014). Nei moronidi, ad esempio, le due vitellogenine VtgAa e VtgAb sembrano legarsi preferenzialmente ai loro recettori specifici Lrp13 e Lr8 (Hiramatsu et al., 2015) (fig.13), mentre nell'*Oncorhynchus* VtgA sembra avere la stessa affinità per entrambi i recettori (Mushirobira et al., 2015).

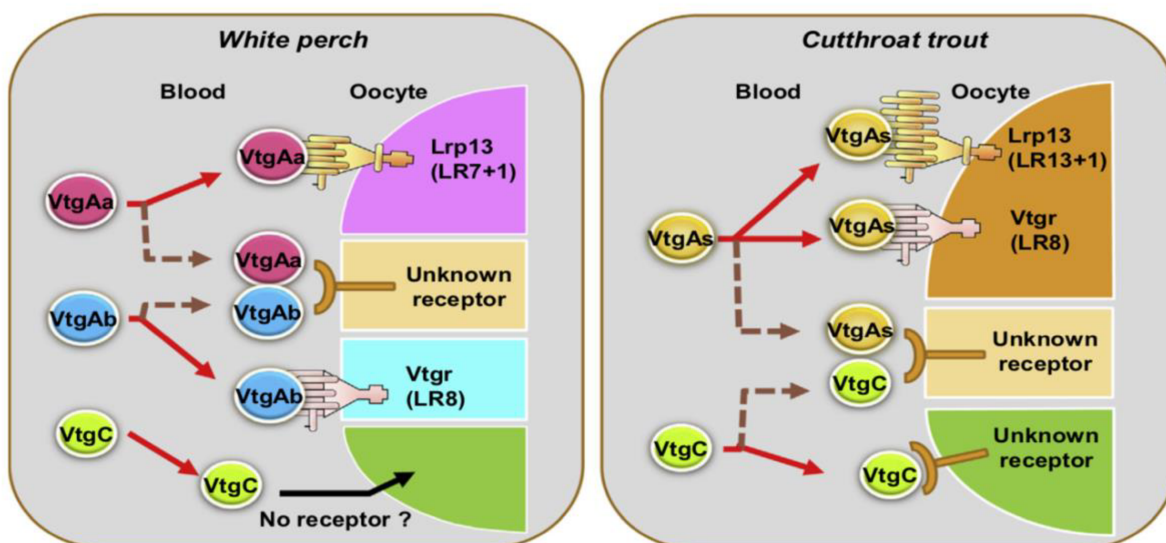


Figura 13. Interazioni tra le differenti Vtgs e i molteplici recettori, figura realizzata da Hiramatsu e colleghi (2015): a destra *O. clarkii* e a sinistra *M.americana*.

Ad oggi, non sono disponibili molte informazioni riguardanti il gene *vtgR* all'interno degli anfibi: la sua sequenza nucleotidica è stata identificata solamente all'interno dello *X. laevis* (Okabayashi et al., 1996).

3. Materiali e metodi

3.1 Campionamento

I campioni utilizzati appartengono ad esemplari di *C. orientalis* che sono stati acquistati dal laboratorio di Filogenesi Molecolare della Facoltà di Scienze, dell'Università Politecnica delle Marche, (Ancona) da rivenditori locali. Gli esemplari utilizzati sono stati sei: tre femmine e tre maschi, acquistati durante la stagione riproduttiva. Tutte le procedure sperimentali sono state approvate dal comitato etico italiano del Ministero della Salute (autorizzazione numero 2E1BD.N.LYB) e sono state seguite tutte le linee guida e i regolamenti pertinenti. Gli esemplari sono stati sacrificati dopo essere stati anestetizzati con MS222 a 2g/l e sono stati prelevati tessuti quali fegato femminile, testicoli e ovaie. Per la conservazione del materiale è stata utilizzata una soluzione di *RNAlater® Tissue Collection* (Applied Biosystems) per la stabilizzazione e la conservazione dell'RNA. Penetrando velocemente in tutto il tessuto questo reagente assicura il mantenimento dell'integrità dell'acido nucleico.

3.2 Estrazione dell'RNA (metodo Trizol)

Per estrarre e utilizzare l'RNA dei tre tessuti, si è seguito un protocollo diviso in cinque fasi che prevede l'utilizzo del TRIzol® (Invitrogen).

1 – *Omogenizzazione del campione*

Si pesano circa 40mg di tessuto da analizzare e lo si immerge in una provetta da 1.5ml contenente 500 μ l di TRIzol. Questa soluzione è una miscela monofasica di fenolo e guanidina isotiocianato che saranno fondamentali per inibire le attività delle RNasi cellulari e quindi per preservare l'integrità dell'RNA. La sua attività risulta essere molto utile nel facilitare l'omogenizzazione del tessuto stesso, andando a disgregare le componenti cellulari. Per l'omogenizzazione si usa un potter mantenendo la provetta in ghiaccio, limitando quindi il surriscaldamento del campione durante l'attività.

2 – Fase di separazione

In questa fase si cerca di separare al meglio le componenti organiche, contenenti anche il DNA, dall'RNA. La provetta contenente l'omogenizzato viene mantenuta per 5 minuti a temperatura ambiente. Successivamente si aggiungono 100 μ l di cloroformio e si inverte vigorosamente la provetta per 15 secondi. Si procede poi con una incubazione a temperatura ambiente di 3 minuti a cui seguono 15 minuti di centrifugazione a 12000xg a 4°C. Al termine della centrifugazione si avrà l'RNA nella fase acquosa, mentre tutto il pellet contenente DNA, lipidi e proteine andrà scartato.

3 – Precipitazione dell'RNA

Si recupera quindi la fase acquosa contenente l'RNA e la si trasferisce in una provetta pulita. Si aggiungono poi 250ml di isopropanolo e delicatamente si

inverte la provetta per 2-3 inversioni laterali. Si incuba la provetta a temperatura ambiente per circa 10 minuti dopo i quali si procede con una centrifugazione di 10 minuti a 12000xg a 8°C. Si ottiene un precipitato traslucido che contiene RNA estratto legato ai sali che verranno eliminati nella successiva fase.

4 – Lavaggio dell'RNA

Si elimina completamente il sovrnatante senza andare a toccare il precipitato ottenuto. Si aggiungono poi 500µl di etanolo al 75% preparato con acqua trattata con DEPC (diethylpirocarbonato). Questo permetterà di sciogliere tutti i sali rimasti nel pellet, liberando l'RNA cellulare. Segue un breve passaggio con il vortex per poi procedere con una centrifugazione di 7500xg per 5 minuti a 8°C. Al termine della centrifugazione si elimina l'etanolo in eccesso e lo si lascia asciugare sotto cappa per un massimo di 3 minuti, mantenendo il tappo della provetta aperto.

5 – Risospensione dell'RNA

L'RNA va poi risospeso in 40µl di acqua RNasi-free e incubato per 10 minuti a 55°C. La provetta può essere poi conservata a -80°C senza pericolo di danneggiamento all'RNA.

3.3 Quantizzazione dell'RNA

È necessario ottenere una quantizzazione dell'RNA estratto prima di procedere con le analisi. Gli strumenti che sono stati utilizzati sono due: il *Nanodrop* e il *Qubit*

1 – *Nanodrop*

È uno spettrofotometro UV-visibile che permette di analizzare la densità ottica di un campione anche in volumi molto esigui (1 μ l) (fig.14). Lo strumento sfrutta la tensione superficiale che si esercita in piccoli volumi quando vengono messi a contatto con due fibre ottiche: si posiziona il volume su una piastra di analisi, la si pone a contatto con la fibra ottica e mantenendo una distanza minima con la piastra di analisi si viene a creare una colonna di liquido che verrà ad essere sfruttata nella lettura. Lo strumento in questo caso è stato utilizzato per quantificare i livelli di RNA andando ad analizzare il rapporto tra la differenza di assorbanza a 260 e a 280nm: si considera un campione con RNA puro se il rapporto A260/280 è compreso fra 1,8 e 2.

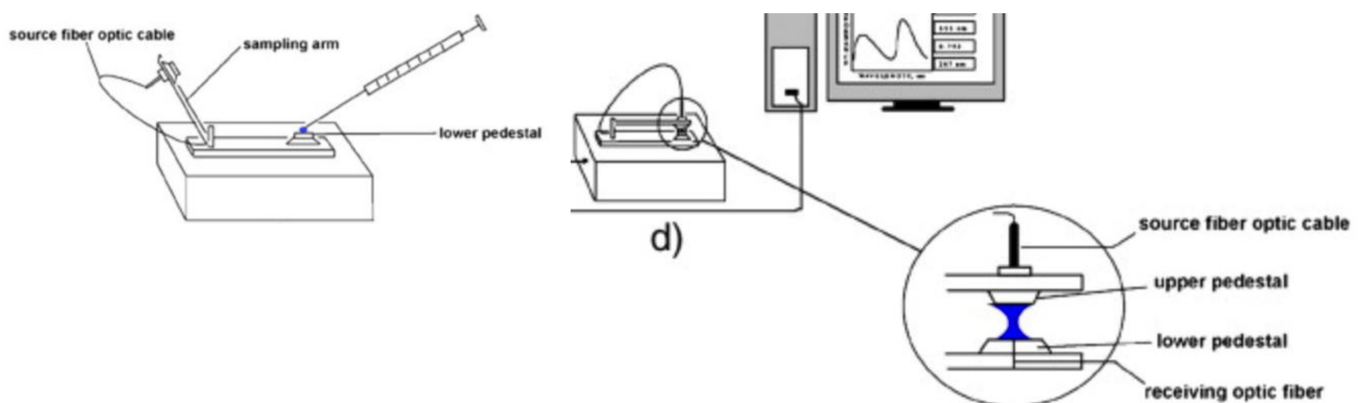


Figura 14. Rappresentazione schematica dell'utilizzo di un Nanodrop realizzata da Andruch e colleghi (2012).

2 – Qubit™ 4 Fluorometer

È uno strumento differente dal Nanodrop perché si basa sulla fluorescenza: il campione viene trattato con una soluzione contenente coloranti fluorescenti capace di legare specificatamente il target di interesse, che può essere RNA, DNA o proteine a seconda delle analisi che si vogliono svolgere con questo strumento. Prima di ogni misurazione si ha la necessità di misurare degli standards per avere una lettura accurata. In particolare, sia il campione che lo standard si preparano con i seguenti volumi:

	Provetta per lo standard	Provetta con campione
Volume dalla Qubit™ working solution	190µl	198µl
Volume dallo standard del kit	10µl	-
Volume del campione	-	2µl
Volume totale in ogni provetta	200µl	200µl

Una volta realizzate le provette, vanno agitate con l'utilizzo di un vortex per circa 5 secondi. Successivamente si incubano le provette a temperatura ambiente per 3 minuti e mantenendo la provetta al riparo da fonti luminose che potrebbero interferire con la sonda della Working Solution. Si effettua poi la lettura dello standard e solo successivamente quella del campione. È importante non oscurare in alcun modo le superfici laterali delle provette con scritte (ad

esempio) per non impedire il passaggio delle radiazioni elettromagnetiche di eccitazione e di emissione del campione. La concentrazione dell'RNA viene poi espressa in quantità di ng/ml.

3.4 Sequenziamento NGS

Il sequenziamento del trascrittoma ottenuto è stato effettuato presso la BGI Tech (Hong Kong), utilizzando la piattaforma Illumina HiSeq2000; le librerie sono state costruite con il kit TruSeq e sottoposte poi a 90 cicli di sequenziamento paired-end. Le sequenze così ottenute sono poi state sottoposte ad un processo di rifinitura mediante l'utilizzo del programma Trimmomatic v.0.39 (Bolger et al.,2014). Il primo passaggio è stato quello di eliminare le sequenze degli adattatori Illumina dalle porzioni di cDNA ottenuti utilizzando l'Illuminaclip. Successivamente si sono analizzate le sequenze tramite una finestra di quattro nucleotidi fatta scorrere dal 5' al 3' e si è valutato il valore della qualità della finestra: sono state rimosse quelle finestre aventi un valore medio di qualità inferiore a 15. Si è effettuato poi un *leading* e *trailing* per eliminare le porzioni a bassa qualità alle due estremità. Le *reads* ad elevata qualità, ottenute al termine di questa procedura, sono state usate per effettuare un *de novo assembly* con il software Trinity v.2.8.3 (Grabherr et al., 2011). usando i parametri di default e impostando la lunghezza minima accettata di

200 nucleotidi. Da questo *assembly* sono poi state eliminati quei *contigs* che sono stati riconosciuti come frammenti di *contigs* più grandi o come mRNAs scarsamente espressi. La qualità finale del trascrittoma *de novo* ottenuto dall'assemblaggio è stata poi valutata con un'analisi mediante software Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simão et al., 2015).

3.5 Ricerca delle sequenze dei recettori della vitellogenina e confronto nelle banche dati

Per identificare quali *contigs* potessero contenere i recettori per Vtg si è pensato di partire dalle sequenze amminoacidiche e nucleotidiche di organismi evolutivamente vicini alla *C. orientalis*. Le sequenze da utilizzare sono state cercate nei database di Ensembl e dell'NCBI: il primo è un browser per genomi di vertebrati che supporta la ricerca in genomica comparativa e gestito dal “European Molecular Biology Laboratory's European Bioinformatics Institute” (EMBL-EBI) che si trova nel Wellcome Genome Campus in Hinxton, nel Regno Unito; il secondo è gestito dal “National Center for Biotechnology Information” (NCBI), curato dalla Library of Medicine (Bethesda USA). La scelta delle sequenze da utilizzare è ricaduta sulla sequenza VtgR di *X. tropicalis* ottenuta da Ensembl (gene ENSXETG00000011709 con transcriptID

ENSXETT00000025606.4). Per la ricerca si è utilizzato BLAST in locale creando un database delle sequenze della *C. orientalis* ottenute dal trascrittoma poi confrontato, mediante la funzione tBLASTn, con la sequenza query VtgR dello *X. tropicalis* selezionata in modo da risalire alle sequenze del trascrittoma che hanno un riscontro con la proteina. La correttezza delle sequenze nucleotidiche ottenute con il BLAST in locale, è stata poi verificata mediante l'analisi delle porzioni codificanti putative mediante il programma l'*Expasy translate tool*, una risorsa bioinformatica online gestita da SIB Swiss Institute of Bioinformatics (www.expasy.org), che permette di identificare il frame di lettura della sequenza nucleotidica, individuando la metionina come primo amminoacido e la presenza di un codone di stop dopo la componente codificante. Oltre al BLAST in locale si è utilizzato anche il programma BLAST (Basic Local Alignment Search Tool) (Altschult *et al.*, 1990) disponibile alla pagina internet ([http:// blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi)) del National Center for Biotechnology Information (NCBI). Ciò ha permesso non solo di valutare se le sequenze ottenute dal BLAST in locale si andassero ad allinearsi con altre *vtgR* o *ldlR* presenti nel database del sito, ma anche di valutare la completezza dell'mRNA in termini di lunghezza, ottenendo altre sequenze proteiche che sono poi state utilizzare per altri studi filogenetici. Infatti, le sequenze ottenute e ritenute buone candidate sono state utilizzate per

la realizzazione di alberi filogenetici preliminari su base amminoacidica insieme a sequenze di altri recettori della famiglia LdlR al fine di confermarne l'ortologia.

3.6 Completamento delle sequenze dei recettori della vitellogenina

Valutando gli allineamenti ottenuti con il BLAST in locale e con il BLAST web server, si è visto che in realtà la probabile sequenza *vtgR* di *C. orientalis* ottenuta dal trascrittoma (*seq1*, TRINITY_DN97501_c0_g5_i1) era mancante dell'estremità al 3'. Si è deciso quindi di completare la sequenza attuando una procedura di laboratorio sperimentale.

3.6.1 Retrotrascrizione

Per ottenere una sequenza completa del *vtgR* è stato necessario effettuare una retrotrascrizione dell'RNA estratto precedentemente. La scelta dell'individuo da utilizzare si è basata sulla quantizzazione dell'RNA estratto e purificato dagli ovari femminili: si è visto che il terzo individuo (F3G) era quello che aveva una quantità maggiore di RNA.

Individuo	Concentrazione RNA (ng/μl)	Volume totale (μl)
F1G	240	3,8

F2G	424	5,7
F3G	813,5	6,0

Il kit utilizzato per effettuare la retrotrascrizione è stato il SuperScript™ III First-Strand Synthesis SuperMix realizzato dalla Invitrogen™. Il suo protocollo permette di sintetizzare il cDNA dall'RNA totale, rilevando anche quelli che hanno una lunghezza di 100bp o oltre i 12kb. Il materiale di partenza può variare da 0,1pg a 5µg per cui in questo caso si è utilizzato 2,5µg di campione, in particolare 3,1µl della soluzione contenente RNA totale. Le soluzioni da utilizzare sono:

- *SuperScript™ III/RNaseOUT™ Enzyme Mix* è la soluzione che contiene la trascrittasi inversa progettata sulla base della M-MLV RT e modificata per ridurre l'attività RNasiH e fornirle una maggior stabilità termica durante la sintesi del cDNA, che può avvenire tra i 45-55°C per aumentare la specificità della reazione. L' *RNaseOUT™*, incluso nella miscela, permette di bloccare eventuali ribonucleasi rimaste in soluzione;
- *2X First-Strand Reaction Mix* è un buffer costituito da 10mM MgCl e 1mM di ogni deossinucleotide (dNTP), utile per la sintesi del cDNA;
- miscele di primer realizzati come esameri random.

La procedura da seguire è la seguente:

1 - si prepara una provetta da 0,2ml contenente le seguenti soluzioni:

Componenti	Volumi (μ l)
RNA purificato	3,1
Primers (50 ng/ μ l)	1
<i>Annealing Buffer</i>	1
<i>RNase/DNase-free water</i>	2,9

2 - si incuba la provetta a 65°C per 5 minuti;

3 - al termine dell'incubazione, la provetta va immersa in ghiaccio per 2 minuti totali; trascorsi 1,5 minuti si aggiungono:

Componenti	Volumi
<i>2X First-Strand Reaction Mix</i>	10 μ l
<i>SuperScriptTM III/RNaseOUTTM Enzyme Mix</i>	2 μ l

4- si procede con la reazione di retrotrascrizione che prevede un passaggio di 10 minuti a 25°C, seguita da 50 minuti a 50°C e termina con 5 minuti a 85°C.

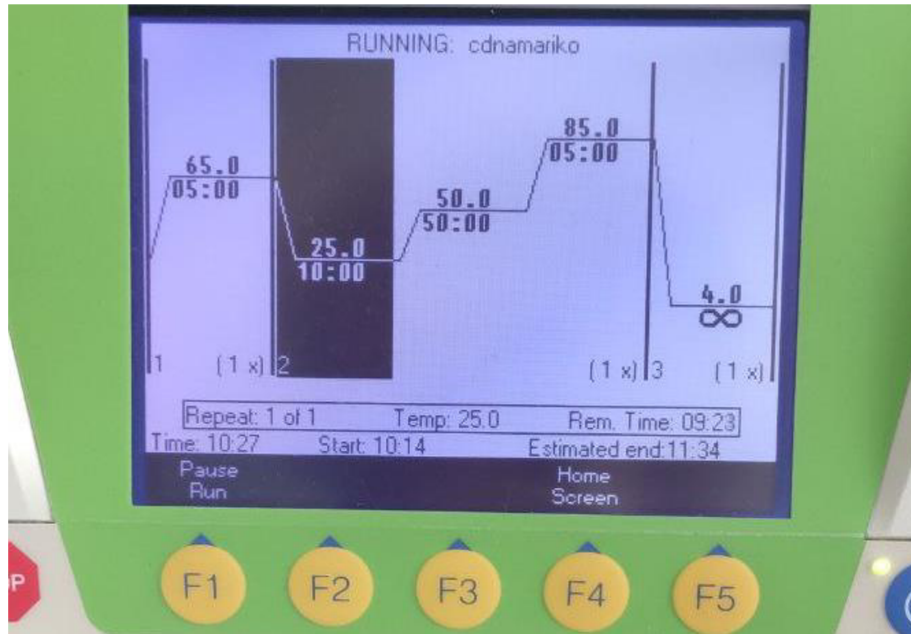


Figura 15. Rappresentazione delle differenti temperature utilizzate per la realizzazione del cDNA.

5- successivamente il campione può essere conservato a -80°C .

3.6.2 Progettazione dei primers

I primers sono stati scelti in base alla reazione a catena della polimerasi (polymerase chain reaction, PCR; Saiki et al, 1988) da effettuare. In particolare, due coppie di primer forward sono state progettate per effettuare una 3'RACE, mentre un ulteriore primer è stato progettato per effettuare un secondo sequenziamento al fine di completare la sequenza.

La sequenza di interesse sulla quale ho progettato i primer per la 3'RACE è la sequenza del trascrittoma della *C. orientalis* identificata come possibile sequenza *vtgR* (*seq1*, TRINITY_DN97501_c0_g5_i1). I primer progettati

come forward sono stati analizzati nelle loro caratteristiche utilizzando il server online del primer-BLAST, fornendo sia la sequenza di interesse che i due primer progettati. Accertata la loro qualità, sono stati ordinati commissionando la realizzazione alla *BMR Genomics*, spin-off dell'Università di Padova (www.bmr-genomics.it). Il terzo primer progettato è stato utile per completare la porzione finale della *seq1* nel sequenziamento. Il primer è stato progettato sulla porzione ottenuta dal primo sequenziamento e ha permesso perciò di completare la sequenza.

3.6.3 Rapid Amplification cDNA ends (RACE)

La *Rapid Amplification cDNA ends* (RACE) è una tecnica che permette di allungare le porzioni incomplete di un cDNA, all'estremità 3' o 5' (Frohman et al., 1988). Durante la realizzazione di un trascrittoma è possibile ottenere delle sequenze incomplete, sia per un mRNA poco abbondante nel tessuto di partenza, sia per un errore nella retrotrascrizione che si traduce come un'impossibilità nell'assemblaggio delle *reads* ottenute. Questa tecnica permette di concentrarsi quindi su un cDNA che non presenta le porzioni iniziali o finali tipiche degli mRNA, sfruttando *primers* specifici in base alla porzione di cui si ha la sequenza.

La 3'RACE prevede due PCR che permettono di amplificare il cDNA incompleto di interesse:

- Con la prima PCR si ottiene un'amplificazione dell'intero frammento di cDNA utilizzando dei primer particolari: il forward disegnato in una porzione conosciuta posta al 5'; il primer reverse è realizzato in modo da avere la sequenza complementare alla coda poli-A presente nel filamento del cDNA, ma, al suo 5', ha una coda che sarà fondamentale per la seconda PCR. Questa coda durante la prima PCR non si appaierà a nessuna sequenza, ma verrà ad essere amplificato a partire dal secondo ciclo. Il prodotto che si ottiene al termine della prima PCR avrà quindi al 5' la sequenza del primer forward, mentre la 3' terminerà con la sequenza dell' adattatore legato al primer reverse.
- Il prodotto della prima PCR viene utilizzato per effettuare una seconda PCR con una coppia di primer differente: il forward sarà più interno del primo, ovvero sarà disegnato in una posizione più al 3' rispetto al forward della prima PCR; il reverse utilizzato sarà invece identico alla coda legata alla sequenza del primer poli-T della prima PCR. Al termine quindi si otterrà un amplificato che contiene tutta l'estremità 3' del cDNA incompleto (fig.16).

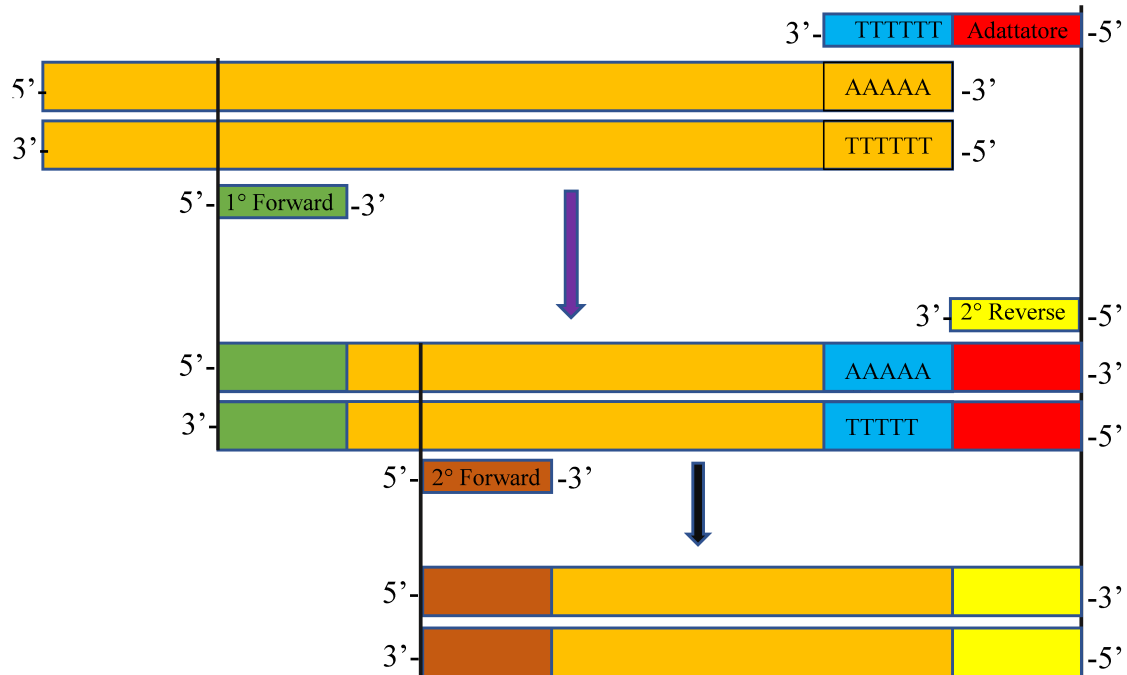


Figura 16. Rappresentazione schematica della 3'RACE: le frecce indicano i prodotti della prima PCR (freccia viola) e della seconda PCR (freccia nera); in ocrea seq1; in verde il primo forward e le porzioni ad esso complementari nel prodotto della prima PCR; in azzurro e rosso il primo reverse utilizzato e le porzioni ad esso complementari nel prodotto della prima PCR; in marrone il secondo forward utilizzato; in giallo il secondo reverse utilizzato. Si noti la differenza di dimensioni tra il primo e il secondo prodotto di PCR realizzato con un secondo forward più interno al primo.

In questa tesi è stata utilizzata la 3'RACE poiché ci si è accorti che la sequenza di interesse mancava della porzione finale al 3'.

Il primo passaggio è stato quello di ottenere il cDNA (paragrafo 3.6.1), per poi procedere con le due PCR:

1°PCR	Forward	3raceA1F	5'-CAAGGCAGTAGGAACAGAG-3'
	Reverse	3race	5'-GACTCGAGTCGACATCGA+Poli-T*-3'
2°PCR	Forward	3raceAF	5'-CTGTGGCGTTGGATGCTG-3'
	Reverse	3raceB	5'-GACTCGAGTCGACATCGA-3'
*Poli-T= TTTTTTTTTTTTTTTTTT			

Prima PCR - Per le reazioni di amplificazione è stato utilizzato il kit *Platinum*[®] *Taq DNA Polymerase* della Thermo Fisher Scientific. Prima di preparare le provette con il cDNA all'interno, si prepara una miscela (*Mix*) in una provetta da 0,5ml con i primers da utilizzare, il *PCR Buffer 10X* (contenente 200mM di Tris-HCl pH 8.4 e 500mM di KCl), i dNTP, il MgCl₂ e la *Taq* polimerasi:

Componenti	Volume (µl)
Primer 3raceA1F (100 µM)	5,91
Primer 3race (300 ng/µM)	3
<i>PCR Buffer 10X</i>	15
dNTP	12
MgCl ₂	4,5
<i>Platinum</i> [®] <i>Taq DNA Polymerase</i>	0,6
Volume totale <i>Mix</i>	41,01

I volumi riportati sono per la realizzazione di tre provette da 0,2ml, ovvero il campione in duplicato(A1a e A1b) e il bianco.

La soluzione viene mantenuta in ghiaccio per conservare l'integrità della *Taq* polimerasi fino al momento dell'utilizzo. Si procede costituendo tre provette da 0,2ml in cui si inserisce la *Mix*, il cDNA (fatta eccezione per la provetta del bianco) e si porta ad un volume finale di 50µl con acqua distillata sterile.

Componenti	Bianco (µl)	A1a (µl)	A1b (µl)
<i>Mix</i>	13,67	13,67	13,67
cDNA	-	1	1
Acqua distillata	36,33	35,33	35,33
Volume finale	50	50	50

Le provette così ottenute vengono poste nel Thermal Cycler MyCycler™ (*BIO-RAD*, Hercules, CA) utilizzando una fase di denaturazione a 94°C per 30 secondi, una fase di annealing a 42°C per 30 secondi e una fase di estensione a 72°C per 1 minuto. Queste fasi vengono ripetute per 30 cicli totali, ma sono precedute da una denaturazione iniziale più spinta con 94°C per 2 minuti e seguite da una fase a 72°C per 10 minuti.

Il prodotto di PCR è stato poi controllato con una corsa elettroforetica sia per visualizzare la presenza dell'amplificato sia per valutare eventuali contaminazioni iniziali. Sono stati caricati 4µl di ogni provetta con 2µl di *Loading Dye* e il gel è stato sottoposto a 70V per 20 minuti.

Seconda PCR – Il protocollo utilizzato nella seconda PCR è lo stesso descritto per la prima reazione di amplificazione ma oltre al campione in duplicato (AFa e AFb) e il bianco è stata costituita una quarta provetta indicata come “*bianco del bianco*” (*BdB*) nel quale viene caricato 1µl del bianco della prima PCR, per essere certi quindi dell'assenza di contaminazione. Il DNA ottenuto dalla prima PCR viene diluito con un rapporto di 1:10 per utilizzarlo nella seconda PCR.

La *Mix* è stata quindi preparata per 4 provette da PCR:

Componenti	Volume (µl)
Primer 3raceAF (100µM)	7,44
Primer 3raceB (300ng/µl)	4
<i>PCR Buffer 10X</i>	20
dNTP	16
MgCl ₂	6
<i>Platinum</i> [®] <i>Taq</i> DNA Polymerase	0,8

Ogni provetta da PCR viene poi preparata secondo questi volumi:

Componenti	Bianco (μ l)	BdB (μ l)	AFa (μ l)	AFb (μ l)
<i>Mix</i>	13,56	13,56	13,56	13,56
cDNA (1:10)	-	-	1	1
Bianco 1°PCR	-	1	-	-
Acqua distillata	36,44	35,44	35,44	35,44
Volume finale	50	50	50	50

Le provette sono poi state inserite nel termociclatore, impostando le stesse temperature utilizzate nella prima PCR. L'unica differenza risiede nel numero di cicli, 15, e nella temperatura della fase di annealing di 52°C, per creare delle condizioni di stringenza maggiori ed evitare l'amplificazione di porzioni di cDNA aspecifico. Il prodotto di PCR è stato poi controllato su gel elettroforetico caricando in ogni pozzetto 4 μ l del prodotto con 2 μ l del *Loading Dye* e fatto migrare a 70V per 20 minuti.

La dimensione attesa del mio amplificato di interesse, calcolato mediante analisi bioinformatiche, è di 1499bp ed è stata verificata mediante ulteriore migrazione elettroforetica utilizzando un marker (*AmpliSize Molecular Ruler* della *BIO-RAD*). Il marker ha la caratteristica di avere bande che vanno da 50bp a 2000bp.

3.6.4 Estrazione e purificazione del prodotto di PCR

Il prodotto di PCR può essere visualizzato con una corsa elettroforetica in TBE che permette di separare gli amplificati secondo dimensione di lunghezza. Il gel utilizzato per estrarre l'amplificato e il tampone da inserire nella camera elettroforetica per la corsa sono stati realizzati con i seguenti volumi:

Gel (agarosio 1,75%) per la corsa elettroforetica	
Acqua deionizzata	24ml
Agarosio	0,525g
TBE 5X	6ml
Colorante di acidi nucleici	2 μ l

Tampone per la corsa elettroforetica	
Acqua deionizzata	200ml
TBE 5X	50ml

Successivamente si è caricato tutto il prodotto di PCR su un nuovo gel realizzato con pozzetti più grandi in modo da contenere tutto il quantitativo ottenuto (circa 50 μ l). Il gel è stato poi sottoposto a una corsa elettroforetica a 70V per 1 ora, in modo da separare eventuali bande multiple.

Il protocollo utilizzato per l'estrazione dell'amplificato dal gel è il *PureLink*® della *Thermo Fisher Scientific* che prevede l'utilizzo di colonnine in silice capaci di trattenere e legare il DNA. Il primo step è quello di tagliare la banda dal gel elettroforetico con un bisturi sterile: si appoggia il gel sul transilluminatore e si effettua il taglio cercando di eliminare la maggior parte del gel che non contiene DNA, accendendo il meno possibile la lampada UV per evitare di indurre la formazione di dimeri di timina. A parte si misura poi il peso iniziale di una provetta da 1,5ml che verrà usata per raccogliere la banda escissa, in modo da calcolare il peso netto della stessa poiché i volumi di buffer da utilizzare sono proporzionali al peso del gel tagliato:

Campione	Banda	Peso provetta (g)	Peso provetta + gel (g)	Δ (g)	3Δ (g)	Volume di L3 (μ l)
1	unica	0,985	1,133	0,147	0,441	440,7
2	unica	0,970	1,085	0,116	0,347	346,8

Quindi si procede con le seguenti fasi:

1 – inserire i tre volumi di *Gel Solubilization Buffer (L3)* nella provetta contenente la porzione di gel. Per sciogliere completamente il gel, si incuba la provetta a 50°C per 15 minuti, invertendo la provetta ogni 3 minuti nei primi 10 minuti;

2 – prelevare tutto il volume ottenuto e caricarlo sulla colonna in silice posizionata in un tubo collettore, fornita dal kit, per permettere la separazione dell'amplificato dalle componenti disciolte del gel. Si effettua poi una centrifugazione a $>12000 \times g$ per 1 minuto. Al termine di questo passaggio si elimina il tampone e il gel disciolto presente nel tubo collettore e si posiziona la colonnina in un nuovo tubo collettore;

3 – aggiungere in colonna 500 μ l di *Wash Buffer (W1)* e con una centrifugazione a $>12000 \times g$ si raccolgono ed eliminano tutte le componenti del gel e del buffer L3 rimasti nella maglia della colonnina. Un'ulteriore centrifugazione alla massima velocità per 2-3 minuti permette di eliminare tutti i residui del buffer utilizzato per i lavaggi;

4 – posizionare la colonnina in una provetta pulita per raccogliere l'amplificato aggiungendo 30 μ l di *Elution Buffer (E1)*. Si incuba a temperatura ambiente per circa 10 minuti, permettendo al *buffer* di raggiungere tutte le porzioni della colonnina. Segue poi una centrifugazione a $>12000 \times g$ per 1 minuto

5– l'amplificato così purificato può essere conservato a 4°C o a -20°C.

Successivamente il DNA ottenuto va precipitato, in modo da eliminare ogni soluzione e rendere il DNA puro. La precipitazione si ottiene inserendo nella provetta 1/10 di sodio acetato e 2 volumi di etanolo assoluto. Si incuba poi

overnight a -20°C . Il giorno successivo si effettua una centrifuga a 4°C a $13000\times g$ per 20 minuti, in modo da precipitare il DNA sul fondo della provetta. Al termine della centrifugazione si elimina tutto il quantitativo di etanolo che potrebbe creare problemi nelle fasi del sequenziamento. In questa fase è importante non danneggiare il DNA precipitato che apparirà come un addensamento translucido. Per far evaporare completamente l'etanolo, la provetta è stata lasciata con il tappo aperto sotto cappa, per 5 minuti. Il DNA poi è stato risospeso in $20\mu\text{l}$ di buffer TE e utilizzato per la quantizzazione.

3.6.5 Sequenziamento di Sanger

Il sequenziamento degli amplificati ottenuti in questa tesi è stato effettuato con il metodo di Sanger presso la *BMR Genomics*, spin-off dell'Università di Padova (www.bmr-genomics.it). I risultati del sequenziamento sono stati forniti tramite elettroferogrammi che sono stati visualizzati e analizzati tramite *FinchTV*, software disponibile al sito <https://digitalworldbiology.com/FinchTV>.

3.7 Analisi dell'espressione genica

3.7.1 Mapping

Le reads generate dalla fase di “ripulitura” o trimming sono state mappate contro la sequenza completata (Seq1) per il recettore della vitellogenina, utilizzata come sequenza di riferimento, mediante l'utilizzo del software CLC Genomics Workbench v.12 (Qiagen). L'allineamento tra la Seq1 e le reads trimmate relative ai tre fegati femminili e alle tre gonadi maschili è stato effettuato utilizzando lo strumento RNA-Seq mapping tool disponibile nel CLC Genomics Workbench impostando i parametri di mapping come altamente stringenti (frazione di lunghezza 0.9 e frazione di similarità 0.9).

3.7.2 Creazione del subset di ortologi per la normalizzazione

Una volta ottenuti i valori relativi al mapping di ciascuna delle reads trimmate contro la nostra sequenza di riferimento, la Seq1, si procede con il calcolo dei valori espressione. Questo step deve essere effettuato tenendo conto di un fattore di correzione calcolato sulla base di un set univoco di geni ortologi a singola copia.

Per la creazione del subset univoco si è proceduto tenendo conto dei risultati dell'analisi BUSCO (Simão et al., 2015). Questo software fornisce una valutazione quantitativa di completezza in termini di contenuto genico atteso

di un assemblaggio genomico, trascrittomico, o di un set di geni annotato: il programma genera dei file output in cui si hanno le attribuzioni dei contigs (nel nostro caso relativi al trascrittoma assemblato de novo) mappati contro dei dataset precompilati, nelle categorie “completo e a singola copia”, “completo e duplicato”, “frammentato”, “BUSCO mancante”.

In linea con queste premesse si prosegue selezionando quei contigs che sono stati attribuiti dal BUSCO alla categoria “completo e a singola copia” per poi procedere al recupero dei valori di “unique gene reads (UGR)” e “Reads Per Kilobase per Million (RPKM)” presenti nei files del mapping dei singoli file trimmati relativi ai tre fegati femminili e alle tre gonadi maschili contro il trascrittoma di riferimento assemblato de novo precedentemente ottenuti in laboratorio dal CLC Genomics Workbench. I valori relativi al totale delle reads mappate (somma delle UGR) ottenuti per ciascun tessuto serviranno poi per il calcolo dei valori di espressione come Trascritti Per Milione (TPM).

3.7.3 Calcolo dei valori di espressione

I livelli di espressione genica sono stati calcolati come trascritti per milione (TPM) in quanto questa metrica permette di comparare efficientemente i livelli di espressione genica sia all'interno che tra i campioni. In questo caso si è

utilizzato un subset di 1,697 orologi e singola copia la cui creazione è stata descritta nel precedente paragrafo.

Inizialmente vengono calcolati i valori di RPKM utilizzando la seguente formula:

$$\text{valore del mapping} / ((\text{lunghezza del gene} / 1000) * (\text{totale delle UGR del set di calibrazione} / 1000000))$$

Per poi procedere con il calcolo dei trascritti per milione:

$$\text{valore degli RPKM} * \text{totale degli RPKM del set di calibrazione} / 1000000$$

3.8 Analisi filogenetica dei dati

L'analisi filogenetica è stata necessaria per valutare la corrispondenza della sequenza di interesse con LdlR di altri organismi e identificarne le distanze evolutive. Le sequenze del trascrittoma della *C. orientalis* identificate come potenziali LdlR e la sequenza VtgR completata, sono state allineate con sequenze prese in analisi nello studio di Reading e colleghi (2014). Le sequenze utilizzate sono state le seguenti:

N°	Classe	Ordine	Specie	LdlR	Accession number (NCBI)
1	<i>Aves</i>	<i>Galliformes</i>	<i>Gallus gallus</i>	VldlR	NP_990560.1
2			<i>Gallus gallus</i>	Lrp1	NP_990573.2

3			<i>Gallus gallus</i>	proLrp1-like	XP_427861.4
4	<i>Mammalia</i>	<i>Primates</i>	<i>Homo sapiens</i>	VldlR	NP_003374.3
5			<i>Homo sapiens</i>	Lrp8	NP_004622.2
6			<i>Homo sapiens</i>	LdlR	NP_000518.1
7			<i>Homo sapiens</i>	proLrp1	NP_002323.2
8			<i>Homo sapiens</i>	Lrp5	NP_002326.2
9			<i>Homo sapiens</i>	Lrp4	NP_002325.2
10			<i>Homo sapiens</i>	Lrp2	NP_004516.2
11			<i>Homo sapiens</i>	Lrp6	AAI17137.1
12			<i>Homo sapiens</i>	Lrp3	NP_002324.2
13			<i>Homo sapiens</i>	Sorla	NP_003096.2
14			<i>Homo sapiens</i>	Lrp11	NP_116221.3
15			<i>Homo sapiens</i>	Lrp10	NP_054764.2
16	<i>Mammalia</i>	<i>Rodentia</i>	<i>Mus musculus</i>	proLrp1	NP_032538.2
17			<i>Mus musculus</i>	LdlR	NP_034830.2
18			<i>Mus musculus</i>	Lrp5	NP_032539.2
19			<i>Mus musculus</i>	Lrp2	NP_001074557.1
20			<i>Mus musculus</i>	Lrp10	NP_075369.2
21			<i>Mus musculus</i>	VldlR	NP_038731.2
22			<i>Mus musculus</i>	Lrp8	NP_001074395.1

23			<i>Mus musculus</i>	Lrp3	NP_001019878.1
24			<i>Mus musculus</i>	Lrp11	NP_766372.1
25	<i>Mammalia</i>	<i>Rodentia</i>	<i>Rattus norvegicus</i>	Lrp6	NP_001101362.1
26			<i>Rattus norvegicus</i>	Lrp4	NP_112612.2
27			<i>Rattus norvegicus</i>	Sorla	NP_445971.1
28	<i>Actinopteri</i>	<i>Cypriniformes</i>	<i>Danio rerio</i>	Lrp6	NP_001128156.1
29			<i>Danio rerio</i>	Lrp4	XP_002663079.2
30			<i>Danio rerio</i>	Lrp2	XP_707254.3
31			<i>Danio rerio</i>	Lrp11	NP_001116346.1
32			<i>Danio rerio</i>	Lrp2	ADK98421.1
33			<i>Danio rerio</i>	Lrp5	NP_001170929.1
34	<i>Actinopteri</i>	<i>Tetraodontiformes</i>	<i>Takifugu rubripes</i>	Lrp1B-like	XP_003974767.1
35	<i>Actinopteri</i>	<i>Perciformes</i>	<i>Morone americana</i>	VtgR	AAO92396.1
36	<i>Actinopteri</i>	<i>Perciformes</i>	<i>Morone saxatilis</i>	VldlR	XP_035513643.1
37	<i>Actinopteri</i>	<i>Salmoniformes</i>	<i>Oncorhynchus clarkii</i>	LdlR	AEO97327.1
38	<i>Actinopteri</i>	<i>Beloniformes</i>	<i>Oryzias latipes</i>	Lrp2	XP_004072920.1

39	<i>Amphibia</i>	<i>Anura</i>	<i>Xenopus laevis</i>	Lrp5	NP_001079163.1
40			<i>Xenopus laevis</i>	VtgR	BAA22145.1
41	<i>Insecta</i>	<i>Blattodea</i>	<i>Periplaneta americana</i>	VtgR	BAC02725.2
42	<i>Insecta</i>	<i>Lepidoptera</i>	<i>Bombyx mori</i>	VtgR	NP_001184180.1
43			<i>Bombyx mori</i>	LpR	NP_001104808.1
44	<i>Insecta</i>	<i>Blattodea</i>	<i>Rhyparobia maderae</i>	LpR	BAE00010.1
45	<i>Amphibia</i>	<i>Caudata</i>	<i>Cynops orientalis</i>	Seq1	TRINITY_DN97501_c0_g5_i1
46			<i>Cynops orientalis</i>	Seq2	TRINITY_DN89930_c0_g2_i1
47			<i>Cynops orientalis</i>	Seq3	TRINITY_DN103341_c0_g1_i1
48			<i>Cynops orientalis</i>	Seq4	TRINITY_DN107861_c0_g1_i1
49			<i>Cynops orientalis</i>	Seq5	TRINITY_DN97073_c0_g2_i1
50			<i>Cynops orientalis</i>	Seq6	TRINITY_DN106116_c0_g1_i1
51			<i>Cynops orientalis</i>	Seq7	TRINITY_DN97073_c0_g3_i1

Il file contenente tutte le sequenze in formato fasta è stato utilizzato per effettuare un multiallineamento attraverso i software Clustal Omega, Kalign, MAFFT e T-Coffee, tutti disponibili su <https://www.ebi.ac.uk> (Madeira et al.,2019):

- Clustal Omega è adatto ad allineamenti medio-grandi (Sievers et al., 2011);
 - Kalign allinea piccole regioni conservate per poi procedere con il resto delle sequenze e risulta essere adatto per lunghi allineamenti (Lassmann et al., 2006);
 - MAFFT, che sfrutta la trasformata di Fourier, è adatto per allineamenti medio-grandi; (Kato et al.,2013)
 - T-coffee, che confronta tutte le tue sequenze due a due, produce un allineamento globale e una serie di allineamenti locali (Notredame et al.,2000).
- Valutando tutti i multiallineamenti, si è scelto di utilizzare quello ottenuto con Kalign poiché risultava essere il migliore; è stato poi realizzato un file NEXUS su cui sono stati annotati il numero di taxa (51 sequenze) e il numero di caratteri per allineamento, considerando sia gli aminoacidi che i gap inseriti (5185) (fig.17).

```

#NEXUS
begin data;
dimensions ntax=51 nchar=5185;
format interleave datatype=protein missing=X gap=-;

matrix

VLDLRGgal_NP_990560.1 -----
VLDLRHsap_NP_003374.3 -----
VLDLRMmus_NP_038731.2 -----
VtgR_Xlae_BAA22145.1 -----
VtgR_Mame_AA092396.1 -----
LRP8_Mmus_NP_001074395.1 -----
LRP8_Hsap_NP_004622.2 -----
LDLR_Ocla_AE097327.1 -----
LDLR_Hsap_NP_000518.1 -----
LDLR_Mmus_NP_034830.2 -----
LipR_Rmad_BAE00010.1 -----
LipR_Bmor_NP_001104808.1 -----
LRP1_Hsap_NP_002323.2 -----
LRP1_Mmus_NP_032538.2 -----
LRP1_Ggal_NP_990573.2 -----
LRP5_Drer_NP_001170929.1 -----
LRP5_Mmus_NP_032539.2 -----
LRP5_Xlae_NP_001079163.1 -----
LRP5_Hsap_NP_002326.2 -----

```

Figura 17. File Nexus realizzato per effettuare analisi filogenetiche.

Per analizzare le relazioni filogenetiche, è stata svolta un'analisi con l'inferenza bayesiana mediante il programma MrBayes (Huelsenbeck et al., 2001), che utilizza l'algoritmo di *Monte Carlo Markov Chain* (MCMC) per valutare gli alberi ottenuti. È un'analisi che sfrutta il metodo di Maximum Likelihood per calcolare la probabilità a posteriori degli alberi filogenetici, ovvero la probabilità che un albero sia corretto, basandosi sulla scelta del più probabile. L'algoritmo MCMC permette di analizzare tutti gli alberi possibili a partire da uno casuale, valutandoli in base alla loro probabilità, scartando quelli con una probabilità minore degli altri. All'aumentare del numero di generazioni effettuate, l'analisi si stabilizzerà su un set di alberi con la stessa probabilità, convergendo quindi verso uno stesso valore. In questo caso, per valutare la

distanza evolutiva considerando le sequenze amminoacidiche, è stato utilizzato il modello “*mixed*” proposto dal programma: si ha un campionamento tra dieci modelli con tassi di sostituzione fissa sfruttando la catena di Markov, in base alla loro probabilità, promuovendo un nuovo modello finale che verrà utilizzato per analizzare le sequenze.

3.9 Predizione di struttura terziaria

Per predire la struttura terziaria di Seq1, si è utilizzato il server I-TASSER (Yang et al.,2015) disponibile on-line (<https://zhanggroup.org/I-TASSER/>). Questo server, partendo da una sequenza amminoacidica, permette di realizzare previsioni di modelli 3D di alta qualità. La realizzazione del modello più idoneo alla sequenza fornita avviene seguendo tre fasi:

- 1) inserendo la sequenza amminoacidica della proteina di cui si vuole predire la struttura terziaria, il server identifica all'interno del database proteico PDB (Berman et al., 2000) delle strutture super-secondarie compatibili con la sequenza fornita, utilizzando il server LOMETS (Wu et al.,2007);
- 2) nella seconda fase, i frammenti identificati nel PDB vengono riassembleati mediante simulazioni Monte Carlo, prendendo in considerazione anche le regioni che non hanno ritrovato un riscontro nel database (solitamente regioni in loop) che vengono ricostruite tramite metodi *ab initio*. Se all'interno del

database non si riscontra alcuna porzione compatibile con la sequenza amminoacidica fornita, l'intera struttura verrà ricalcolata con metodi *ab initio*. In questa fase vengono identificati anche gli stadi a minor energia tramite l'algoritmo SPICKER (Zhang et al., 2004);

3) nella terza fase si rielabora l'intero modello partendo dai *centroids* identificati da SPICKER (Zhang et al., 2004) , tenendo in considerazione le informazioni di ingombro sterico identificate da LOMETS (Wu et al., 2007) e valutate da TM-align (Zhang et al., 2005). Questa terza fase è utile per perfezionare i modelli ottenuti, scartando i ripiegamenti caratterizzati da un elevato ingombro sterico o da alta energia: solo le strutture a più bassa energia vengono mantenute. Il modello finale viene poi ottimizzato in base ai suoi legami idrogeno tramite REMO (Yunqi et al., 2009).

4. Risultati

4.1 Ottenimento delle sequenze del recettore della vitellogenina dal database

L'identificazione della sequenza di *vtgR* nel trascrittoma della *C. orientalis* è avvenuta utilizzando inizialmente il BLASTN in locale, impostando come query *vtgR* dello *X. laevis*, ottenuta dal database nucleotidico dell'NCBI (AB006906.1). I risultati non sono stati utili per la ricerca, dato che l'allineamento avviene solo per porzioni limitate e con un punteggio piuttosto basso (78.7 score)(fig.18).

```
BLASTN 2.11.0+

Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb
Miller (2000), "A greedy algorithm for aligning DNA sequences", J
Comput Biol 2000; 7(1-2):203-14.

Database: cynops_assembly_HQ.fasta
45,831 sequences; 73,056,960 total letters.

Query= AB006906.1 Xenopus laevis mRNA for vitellogenin receptor, complete
cds
Length=3566

Sequences producing significant alignments:
Score   E
(Bits)  Value
TRINITY_DN108540_c1_q1_i3     78.7  5e-13
TRINITY_DN108540_c1_q1_i6     78.7  5e-13

>TRINITY_DN108540_c1_q1_i3
Length=522
Score = 78.7 bits (42), Expect = 5e-13
Identities = 119/153 (78%), Gaps = 17/153 (11%)
Strand=Plus/Minus
Query 3144 AGTACTGTACAAAGCACATTTCCATAGAAAGCCATATCCAG-AA--A--TT-GTGCTGTA 3197
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct 399  AGTACTGT--AAAGCACATTTCCATAGAAAGCCATATCCAGCAATGAAATTTGTGCT--A 344

Query 3198 TGT-AGT-A-AGTTATCCTGTACATATGTGTAAGGCAGTTTGTACATATCCTTAGAA-A 3253
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct 343  TGTTAGTGACACCCACCTGTACATATTTGTATAGGCCATGTGTAATAT--TTCGAAGA 286

Query 3254 GAACACAATCACAATCTGAAGCACTTTGAAAT 3286
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct 285  GATG--CAATCACTATCTTAAAGCACTTTGAAAT 254
```

Figura 18. Risultato della prima ricerca in BLASTN, effettuata in locale, ottenuto considerando come database il trascrittoma della *C. orientalis* e come query la sequenza *vtgR* dello *X. laevis*.

La ricerca a questo punto si è incentrata sull'utilizzo di una sequenza proteica di un organismo evolutivamente vicino alla *C. orientalis*, ma differente dallo

X. laevis vista la scarsità dei risultati: dal database di Ensembl (www.ensembl.org) si è ottenuta la sequenza amminoacidica di Vldlr dello *X. tropicalis* (gene ENSXETG00000011709.4 con il trascritto ENSXETT00000025606.4) e si è effettuata una ulteriore ricerca in TBLASTN in locale (fig.19).

```
TBLASTN 2.11.0+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A.
Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J.
Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of
protein database search programs", Nucleic Acids Res. 25:3389-3402.

Database: cynops_assembly_H0.fasta
45,831 sequences; 73,056,960 total letters.

Query= ENSXETG00000011709:ENSXETT00000025606.4 peptide: ENSXETP00000025606
pep:protein_coding

Length=878

Sequences producing significant alignments:
```

	Score (Bits)	E Value
TRINITY_DN97501_c0_g5_i1	786	0.0
TRINITY_DN89930_c0_g2_i1	779	0.0
TRINITY_DN103341_c0_g1_i1	751	0.0
TRINITY_DN107861_c0_g1_i1	435	7e-131
TRINITY_DN97073_c0_g2_i1	412	4e-123
TRINITY_DN106116_c0_g1_i1	361	7e-106
TRINITY_DN106116_c0_g1_i5	361	7e-106
TRINITY_DN97073_c0_g3_i1	325	1e-100
TRINITY_DN102921_c0_g1_i3	230	3e-62
TRINITY_DN102921_c0_g1_i2	229	1e-61
TRINITY_DN106918_c0_g1_i1	211	8e-56
TRINITY_DN97501_c0_g3_i1	170	1e-49
TRINITY_DN97501_c0_g4_i1	150	2e-42
TRINITY_DN107620_c3_g1_i2	161	6e-40
TRINITY_DN104638_c0_g1_i1	131	1e-30
TRINITY_DN105278_c0_g2_i1	127	1e-29
TRINITY_DN106762_c0_g1_i1	117	5e-26
TRINITY_DN106762_c0_g1_i2	117	5e-26
TRINITY_DN170109_c0_g1_i1	88.2	6e-21
TRINITY_DN104410_c0_g3_i3	97.1	7e-21
TRINITY_DN74405_c0_g1_i1	86.7	7e-20
TRINITY_DN100167_c0_g1_i1	94.7	3e-19
TRINITY_DN102679_c0_g3_i3	88.6	3e-17
TRINITY_DN99927_c0_g2_i5	80.1	8e-15
TRINITY_DN106907_c0_g1_i1	79.0	2e-14
TRINITY_DN105718_c0_g1_i2	72.4	2e-12

Figura 19. Risultati del TBLASTN effettuato con la query dello *X. tropicalis*.

I risultati vengono presentati sulla base del valore dello score (dal più alto al più basso) e dell'E-value, una misura che indica l'affidabilità del punteggio di score, ed identificano diverse porzioni del trascrittoma che hanno un riscontro

con la query dello *X. tropicalis*. Analizzando le subject ottenute, sono sette sequenze a mostrare il miglior allineamento con le porzioni della query e a presentare un punteggio di score molto elevato (>300).

Sequenza nel trascrittoma	Nome assegnato	lunghezza	frame	score
TRINITY_DN97501_c0_g5_i1	Seq1	2040	-1	786
TRINITY_DN89930_c0_g2_i1	Seq2	3084	-3	779
TRINITY_DN103341_c0_g1_i1	Seq3	6086	-1	751
TRINITY_DN107861_c0_g1_i1	Seq4	14632	-3	435
TRINITY_DN97073_c0_g2_i1	Seq5	10142	-1	412
TRINITY_DN106116_c0_g1_i1	Seq6	6680	-3	361
TRINITY_DN97073_c0_g3_i1	Seq7	3710	-2	325

Le sequenze nucleotidiche corrispondenti alle subject sono poi state tradotte in proteina con il programma Translate tool dell'Expasy e utilizzate per effettuare un BLASTP al fine di verificare la corrispondenza con LdlR di altre specie: tutte le sequenze prese in considerazione mostravano somiglianza con recettori della famiglia LdlR di differenti organismi, ma l'unica ad avere una corrispondenza con VldlR è stata la Seq1. È stata anche realizzata un'analisi filogenetica basata su tutte le sequenze identificate nel trascrittoma come

possibili candidate e LdlR di differenti organismi appartenenti a diverse classi tassonomiche. Le sequenze utilizzate in questa fase sono state le stesse prese in considerazione dallo studio di Reading e colleghi, 2014. Per effettuare una prima analisi filogenetica si è utilizzato il programma MegaX (Stecher et al., 2020) applicando il metodo della Massima Verosimiglianza (Maximum Likelihood). I risultati (fig.20) hanno evidenziato le relazioni tra le sequenze identificate e le sequenze della famiglia LdlR: Seq1 è quella più vicina al VtgR dello *X. laevis* ed è l'unica sequenza che raggruppa con le VldlR prese in analisi; Seq2 è simile alle LdlR; Seq3 raggruppa con le Lrp8; Seq4 con le Lrp2; Seq5 e Seq7 raggruppano con le Lrp1 mentre Seq6 con i recettori identificati come Lrx+1. Si è deciso quindi di proseguire le analisi considerando principalmente Seq1 nonostante dagli allineamenti risultasse incompleta nella porzione 3'. La lunghezza della porzione mancante, circa 1300bp , è stata determinata mediante un allineamento con la sequenza VtgR di *X. laevis* (BC070552.1)



Figura 20. Albero filogenetico delle sequenze prese in esame da Reading e colleghi (2014) ottenuto con MEGAX applicando il metodo di Maximum Likelihood, e delle sequenze individuate nel database del trascrittoma della *C. orientalis* sono indicate dalle frecce rosse.

4.2 Progettazione dei primers

Poiché è stata necessaria una 3'RACE per completare Seq1, sono stati progettati di primers forward sulla sequenza della TRINITY_DN97501_c0_g5_i1 scritta in direzione 5'-3' sono stati progettati i due primers forwards.

```
>TRINITY_DN97501_c0_g5_i1
TAACACGCATCTCCCCGCCACCTCCTTCTCGGGCTTGCATCAAAGAGCAGACGGCGCTCACCTGA
CGCCGACCGGCCGCTTGCAGACACTCAGGCGCACGTGCAGCGCGTGCTCTCTGATTGCACCTAGGC
TCGGCTGGGGGAATTGGGATCGGCTGAGTTGCGCGCGCTCTGGTGGCAGCGCGCGGCTGCGCAAG
CGGCAGGGGGCGCCGGAGAGCGCTATAAGGAGCGCGCGCCCTCCAGGACTGCTTCTATTGCAC
CTTCCCGGCCACGGAGACAGGCGGCGTCTGGTTCAGCGAGCCCGGTGGAGAGGGAGCAGAGCCGGC
TGTGCGGGGTTGGAGCTTGCATCAGTCTTCTTTGGACCCGTGCCTGAATTTGTATTGCTGGATCGG
GTTGTTGGATCTTCTGAACCCCTGGAACAGGGGCTGTTGGCTGCGGAGATGTCGCTGCGGCGCGGA
GGTCTGGGCGCTGTGCTGCTGGCGCTGCTGCCGTTGCTGCTGGCCGCCAGCCGGCCTTCGTGGGGT
TCAAGAACAAAATGTGAAGACAATCAATTCCAGTGCCGCAATGGACGTTGTATAACATCTGTCTGG
CAGTGTGATGGGGATGACGATTGTTTCCAGATGGTACAGATGAGCTATCTTGTGTTAAGAAGACTTGT
GCGGAGTCTGACTTTGTATGCAGTAATGGGCAGTGTGTTTCTGAAAGGTGGCAGTGTGATGGAGAC
CAAGACTGTGAAGATGGATCTGACGAGCATCCAGAGTTCTGTCAATTTAAGAAGACTTGCCGGGTGACT
GAAATCAGTTGTGGTCCACGATCAACGCAGTGTATTCCAGTTTCTTGGAGCTGTGATGGAGAGTCT
GATTGTGACAGTGGAGAGGATGAAACCAATTGTGGCAACATTACCTGCAGTCCAGATGAGTTCACG
TGTTCCAGCGGGCGATGCGTTTCTAAGATATTTTTCTGCAACGGCCATGACGACTGCGGCGATGGA
AGTGACGAGCTGGACTGCACCCCTCCGACCTGCGGGGCTCACGAGTTTCAGTGCAAGAAGTCTTCC
TGCATCCCCATCAGCTGGGTGTGTGATGACGATGCAGACTGCGCGGACCCTCGGATGAGTCGATG
GAGCAGTGTGGGCGCCAGCCGGCCTCCCCGCAGCGCTGCCCGGAGAGCGAGATCCAGTGTGGCTCT
GGGGAGTGCATTCACAAGAGGTGGCGCTGCGATGGCGACGCAGACTGCAAAGACAAAAGCGATGAG
ATTAAGTGTCTCCTCGTACTTGTGCGACCAGACCAATTCAAATGTCAAGATGGGAACTGTGTTTCA
GGCAGCCGGCAGTGTAAATGGTGTAAAGAGACTGTCTTGATGGCACTGACGAAATAAACTGTAAAAC
GTCAATCAGTGTCTCAGGACCTGGTAAATTCAGTGTAGAAGTGGGGAATGCATTGAAATAAGCAAG
GTTTGTAAATGAGCTACAGGACTGCAGAGACTGGAGTGTGAACCGCTGAAGGAATGTGGTCTAAAT
GAATGCCTGGTAAATAATGGCGGCTGCTCCCACATCTGCAGAGACTTGATCGTGGGCTATGAATGT
GAATGCCCTGCCGGGTTTGGAGCTCATTGACAGGAAAACATGCGGAGATATTGATGAATGTCACAAT
```


CCTGGAATCTGCAGTCAAATTTGTGTGAACTTGAAAGGCGGCTACAAGTGTGAATGCAGCCGTGGG
 TATCAAATGGATCTTGCCACTGGAGTGTG **CAAGGCAGTAGGAACAGAG**CCATGTTTAATCTTCACA
 AACCGAAGAGACATTAGAAAAGATTGGCCTGGAAAGGAAGGAGTACAAACAGTTAGTGGAAACAATTG
 AGGAACA **CTGTGGCGTTGGATGCTG**ACATTGCTGGCCAGAATATTTTTTTGGGCTGACGTAAACCAA
 AAATCAATTTTTTCGTGCATCAGTGGATTCCCGAGACAAAATTGGCCGCCATGTGAAAATCATAAAC
 AATTTGCACAGCCCTGCATCAATTGCTGTAGACTGGGTTTATAAGAATATCTATTGGGCA

In **verde** la porzione del primer 3raceA1F;
 in **azzurro** la porzione del primer 3raceAF.

Per valutarne le caratteristiche come lunghezza, *self complementarity*, *self complementarity* al 3' e la temperatura di melting (Tm), i due primers sono stati inseriti nel server del primer-BLAST (Ye et al., 2012) (fig.21-22) insieme a *seq1* incompleta: questo passaggio ha permesso anche di confermare l'assenza di altre porzioni complementari ai primer ipotizzati.

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CAAGGCAGTAGGAACAGAG	Plus	19	1746	1764	54.92	52.63	2.00	0.00

Figura 21. Caratteristiche del primer 3raceA1F identificate con primer-BLAST.

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CTGTGGCGTTGGATGCTG	Plus	18	1856	1873	59.13	61.11	2.00	1.00

Figura 22. Caratteristiche del primer 3raceAF identificate con primer-BLAST.

Come reverse sono stati utilizzati due particolari primers: il primo utilizzato è composto da un adattatore a cui è stato legato al 3' una coda poli-T, utile per riconoscere la porzione della coda poli-A; il secondo è un primer che risulta essere complementare alla sequenza dell'adattatore del primo primer.

Forward prima PCR - 3'RACE	
Nome	3raceA1F

Sequenza	CAAGGCAGTAGGAACAGAG
Lunghezza	19bp
Peso molecolare	5920,8µg/µmole
Coeff.di estinzione	236,5 OD/µmol
Tm	55 °C
%GC	53
N° di moli	36,8 nmol

Reverse prima PCR - 3'RACE	
Nome	3race
Sequenza	GACTCGAGTCGACATCGA+Poli-T*
Lunghezza	35 bp
Peso molecolare	10680 µg/µmole
Coeff.di estinzione	359,0 OD/µmol
%GC	28,6
N° di moli	45,3 nmol

*Poli-T= TTTTTTTTTTTTTTTTTT

Forward seconda PCR - 3'RACE	
Nome	3raceAF
Sequenza	CTGTGGCGTTGGATGCTG
Lunghezza	18 bp
Peso molecolare	5578,6 µg/µmole

Coeff.di estinzione	187,7 OD/ μ mol
Tm	59 °C
%GC	61
N° di moli	29,9 nmol

Reverse seconda PCR - 3'RACE	
Nome	3raceB
Sequenza	GACTCGAGTCGACATCGA
Lunghezza	17 bp
Peso molecolare	5195 μ g/ μ mole
Coeff.di estinzione	185,0 OD/ μ mol
Tm	55,2 °C
%GC	58,8
N° di moli	59,6 nmol

4.3 Completamento della sequenza seq1

Per completare la sequenza di *seq1*, è stata svolta una 3'RACE. Prima di procedere con questa tecnica, è stato necessario ottenere il cDNA a partire dall'RNA isolato dal tessuto ovarico della terza femmina della *C. orientalis* (F3G) secondo le modalità descritte in precedenza. Successivamente sono state svolte le due PCR utilizzando la coppia di primers progettati in base alla porzione già conosciuta di Seq1.

1°PCR	Forward	3raceA1F	5'-CAAGGCAGTAGGAACAGAG-3'
-------	---------	----------	---------------------------

	Reverse	3race	5'-GACTCGAGTCGACATCGA+Poli-T*-3'
2°PCR	Forward	3raceAF	5'-CTGTGGCGTTGGATGCTG-3'
	Reverse	3raceB	5'-GACTCGAGTCGACATCGA-3'
*Poli-T= TTTTTTTTTTTTTTTTTT			

In entrambe le PCR, l'amplificato è stato controllato e valutato tramite una corsa elettroforetica:

- della prima PCR sono stati caricati su gel elettroforetico i due campioni A1a e A1b (amplificati in duplicato per avere più materiale per il sequenziamento) e il rispettivo bianco (B)(fig.23) .

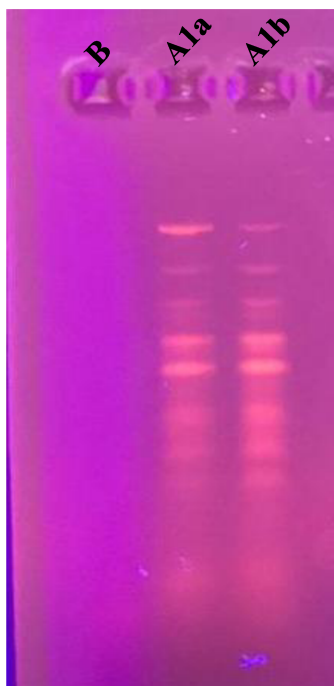


Figura 23. Gel elettroforetico della prima PCR al transilluminatore: si nota l'assenza di contaminazioni nel bianco (B) e gli amplificati identici nei due campioni (A1a e A1b).

Al termine della corsa elettroforetica, il bianco è risultato privo di contaminazioni poiché non si sono presentate bande di amplificati aspecifici. Nei pozzetti dove sono stati caricati i due amplificati invece, si è osservata la presenza di più bande, identiche tra i due campioni, indice del fatto che la coppia di primer utilizzati ha amplificato anche altre porzioni del trascrittoma oltre alla sequenza di interesse.

- effettuata poi anche la seconda PCR con la seconda coppia di primers, si sono controllati i risultati effettuando una nuova corsa elettroforetica, caricando non solo il bianco (B) e i due campioni in duplicato (AFa, AFb) ma anche un quarto campione realizzato con il bianco della prima PCR per avere la certezza dell'assenza di contaminazioni (fig.24).



Figura 24. Gel elettroforetico della seconda PCR al transilluminatore: si nota l'assenza di contaminazioni nel bianco (B) e nel bianco del bianco (BdB), mentre i campioni(AFa e AFb) hanno gli identici amplificati.

Osservando il gel al transilluminatore si nota una banda molto marcata: con la seconda coppia di primers, avendo aumentato le condizioni di stringenza della reazione, si è amplificata la porzione di cDNA di interesse, eliminando gli aspecifici ottenuti nella prima PCR. Poiché erano presenti due bande di diversa intensità, per identificare con certezza l'amplificato corrispondente a *seq1* si è effettuata una seconda corsa elettroforetica prendendo in considerazione anche un marker di lunghezza. La banda con maggior intensità è risultata essere di circa 1500bp, lunghezza compatibile con quella stimata della porzione mancata di *seq1*.

Il risultato del sequenziamento di *seq1* viene fornito come un elettroferogramma che è possibile analizzare con il software FinchTV (www.geospiza.com/finchtv) (fig.25).

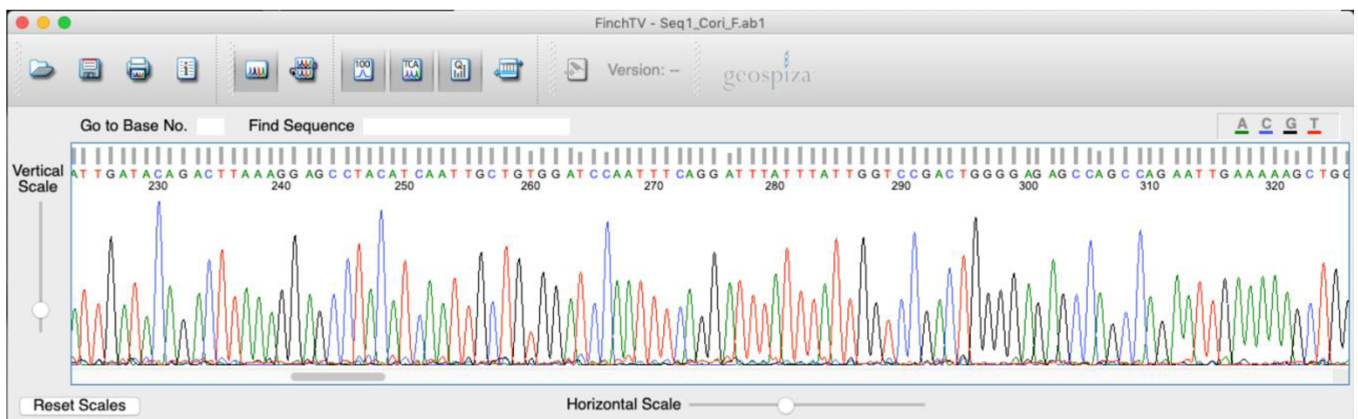


Figura 25. Elettroferogramma ottenuto dal primo sequenziamento: in verde l'adenina, in blu la citosina, in nero la guanina e in rosso la timina .

Un primo sequenziamento, svolto utilizzando il secondo primer forward della 3'RACE (3raceAF), ha permesso di completare con certezza solo la prima

porzione mancante di *seq1*. Poiché la lunghezza della porzione da completare (circa 1449bp) impedisce di ottenere un sequenziamento attendibile con una sola reazione, è stato svolto un secondo sequenziamento.

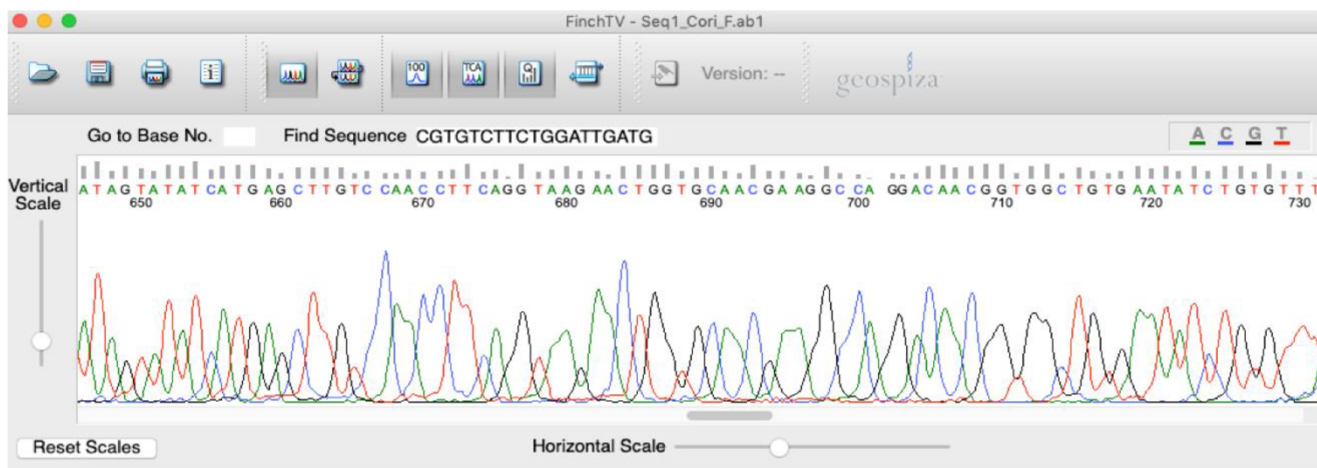


Figura 26. Parte finale della porzione attendibile del primo sequenziamento: la presenza di doppi picchi e la diminuzione della certezza nella lettura dei singoli nucleotidi rendono necessario un secondo sequenziamento.

La seconda reazione è stata realizzata con un secondo primer progettato al termine della prima porzione completata: la presenza di doppi picchi e l'aumento del rumore di fondo rende affidabile il primo sequenziamento solo entro i primi 580-600bp, dove appunto è stato progettato il secondo primer da utilizzare (C1F). La valutazione è stata svolta utilizzando il primer-BLAST (fig.27):

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	ATCGTGTCTTCTGGATTGATG	Plus	21	454	474	55.78	42.86	3.00	0.00

Figura 27. Caratteristiche del primer C1F identificate con primer-BLAST

Primer per sequenziamento	
Nome	C1F
Sequenza	ATCGTGTCTTCTGGATTGATG
Lunghezza	21 bp
Peso molecolare	6459,2 $\mu\text{g}/\mu\text{mole}$
Coeff.di estinzione	222,6 OD/ μmol
Tm	58°C
%GC	43
N° di moli	18 nmol

La sequenza ottenuta è stata poi unita alla porzione già conosciuta e tradotta in proteina con il *Translate tool* per identificare sia la porzione codificante (dal nucleotide 444 al 3051).

La sequenza completata risulta essere di 3332 nucleotidi:

TAACACGCATCTCCCCGCCACCTCCTTCCCTCGGGCTTGCATCAAAGAGCAGACGGCGCTCACCTGACGCGACCGGCCGCTTGCAGACACTCAGGCGCACGTGCAGCGCGTGCTCTCTGATTGCACCTAGGCTCGGCTGGGGGAATTGGGATCGGCTGAGTTGCGCGCGCGTCTGGTGGCAGCGCGCGGCTGCGCAAGCGGCAGGGGGCGCCGGAGAGCGCTATAAGGAGCGCGCGCCCTCCAGGACTGCTTCCTATTGCACCTTCCCGCCACGGAGACAGGCGCGTCTGGTTCAGCGAGCCCGGTGGAGAGGGAGCAGAGCCGGCTGTGCGGGGTTGGAGCTTGCATCAGTCTTCTTTGGACCCGTGCCTGAATTTGTATTGCTGGATCGGGTTGTTGGATCTTCTGAACCCCTGGAACAGGGGCTGTTGGCTGCGGAGATGTCGCTGCGGCGCGGAGGTCTGGGCGCTGTGCTGCTGGCGCTGCTGCCGTTGCTGCTGGCCGCCAGCCGGCCTTCGTGGGGTCAAGAACAAAATGTGAAGACAATCAATTCCAGTGCCGCAATGGACGTTGTATAACATCTGTCTGGCAGTGTGATGGGGATGACGATTGTTTCAGATGGTACAGATGAGCTATCTTGTGTTAAGAAGACTTGTGCGGAGTCTGACTTTGTATGCAGTAATGGGCAGTGTGTTCCTGAAAGGTGGCAGTGTGATGGAGACCAAGACTGTGAAGATGGATCTGACGAGCATCCAGAGTTCGTGTCATTTAAGAAGACTTGCCGGGTGACTGAAATCAGTTGTGGTCCACGATCAACGCAGTGTATTCCAGTTTCTTGGAGCTGTGATGGAGAGTCT

GATTGTGACAGTGGAGAGGATGAAACCAATTGTGGCAACATTACCTGCAGTCCAGATGAGTTCACG
TGTTCCAGCGGGCGATGCGTTTCTAAGATATTTTTCTGCAACGGCCATGACGACTGCGGCGATGGA
AGTGACGAGCTGGACTGCACCCCTCCGACCTGCGGGCTCACGAGTTTCAGTGCAAGAACTCTTCC
TGCATCCCCATCAGCTGGGTGTGTGATGACGATGCAGACTGCGCGGACCACTCGGATGAGTCGATG
GAGCAGTGTGGGCGCCAGCCGGCCTCCCCGAGCGCTGCCGGAGAGCGAGATCCAGTGTGGCTCT
GGGAGTGCATTCACAAGAGGTGGCGCTGCGATGGCGACGCAGACTGCAAAGACAAAAGCGATGAG
ATTAAGTGTCTCCTCGTACTTGTTCGACCAGACCAATTCAAATGTCAAGATGGGAACTGTGTTTCA
GGCAGCCGGCAGTGTAAATGGTGTAAAGAGACTGTCTTGATGGCACTGACGAAATAAACTGTAAACT
GTCAATCAGTGCTCAGGACCTGGTAAATTCAAGTGTAGAAGTGGGGAATGCATTGAAATAAGCAAG
GTTTGTAAATGAGCTACAGGACTGCAGAGACTGGAGTGTGAACCGCTGAAGGAATGTGGTCTAAAT
GAATGCCTGGTAAATAATGGCGGCTGCTCCCACATCTGCAGAGACTTGATCGTGGGCTATGAATGT
GAATGCCCTGCCGGTTTGGAGCTCATTGACAGGAAAACATGCGGAGATATTGATGAATGTCACAAT
CCTGGAATCTGCAGTCAAATTTGTGTGAACTTGAAAGGCGGCTACAAGTGTGAATGCAGCCGTGGG
TATCAAATGGATCTTGCCACTGGAGTGTGCAAGGCAGTAGGAACAGAGCCATGTTTAACTTTCACA
AACCGAAGAGACATTAGAAAGATTGGCCTGAAAGGAAGGAGTACAAACAGTTAGTGAACAATTG
AGGAACA**CTGTGGCGTTGGATGCTG**ACATTGCTGGCCAGAATATTTTTTGGGCTGACGTAAACCAA
AAATCAATTTTTCGTGCATCAGTGGATTCCCGAGACAAAATTGGCCGCCATGTGAAAATCATAAAC
AATTTGCACAGCCCTGCATCAATTGCTGTAGACTGGGTTTATAAGAATATCTATTGGGCAGACACC
GGTTTAAAGACTATCTCTGTATCAAACCTTAATGGCTCACGGACAAAAATATTGATTGATACAGAC
TTAAAGGAGCCTACATCAATTGCTGTGGATCCAATTTTCCAGATTTATTTATTGGTCCGACTGGGGA
GAGCCAGCCAGAATTGAAAAAGCTGGAATGAATGGAATCGACAGGCAAAAGTTGGTGACCACAGAC
ATTCAGTGGCCAAATGGAATTGCACTAGACCTTGTA AAAAGTCGCCTCTACTGGCTGGATTCCAAG
CTACACATGCTGTCCAGTGTGGACTTTAATGGCCAGGATCGCAGAATAGTCTGAAATCCTTGGAG
TTCTTTGCTCACCCCTCTGGCTCTTACGATATTTGAGG**ATCGTGTCTTCTGGATTGATG**GTGAAAAT
GATGCAATTTATGGTGCCAATAAATTTACCGCCAAGAAATAGCAACGCTAGTTAACAACCTAATT
GATGCCCAGGATATTATAGTATATCATGAGCTTGTCCAACCTTCAGGTAAGAACTGGTGAACGAA
GGCCAGGACAACGGTGGCTGTGAATATCTGTGTTTACCTGCTCCACAGATAAATAGTCATTCCCC
AAGTACACATGTGCATGTCTGCTGGATCTGAACTGGAGGATGGTGGTGAAGATGTACAAAAGAT
GTTGGAGGTACTGCAGCTTATGGTGAACAAAAGATGCCAGTACACGAGGGAAATCTCCGATCGAA
GAACACAGCTCCTCGGTATTTGAAGCAAATCAGTCTACAAAAGGATCTGCAGCCTGGGCTATTCTT
CCTCTTTTGCTACTTGCAGTGGCAGCTGGTGGTGGTTACCTGGTGTGGCGAAATTGGAAGCGCAAA
AATATGAAAAGTATGAACTTTGACAATCCAGTGTACTTGAAGACCACAGAGGAGGACCTCAACATA
GATCTAGGTGACACAGTACTTCAAGTAGGACACACATATCCAGCAATATCTATTGTGAGCACAGAA
GATGACCTGGCG**TGAG**GCATAGGCTGGTCCGGTCGGCATTACCTCCCTCCTTCAGTATACCGCGCTT
GTACAGTTTTTTCTTTTTCTGAATGGTAACCAAGTCAGCAACTGAACACATCCCGCCTTGAAGAA

TGCGAAGAGAGCCTTTTGTATGCGTGAAACACATTTTGGGTGCAATTGTTTTAAATGCTTACAAAT
 GAAACCTCTTGTAATATTTCTTAACCGTAGAGTTCAGCTTGTGATGTGGTTACCATTTAGTCTCC
 ACCCTTTGGACTTTGT CAGTATTA AAAAAAAAAA

In grigio le porzioni non codificanti;
 in verde il codone di inizio;
 in azzurro la porzione del primer 3aceAF;
 in giallo la porzione del primer C1F;
 in rosso il codone di stop.

La traduzione della sequenza mostra una proteina di 868 amminoacidi:

MSLRRGGLGAVLLALLPLLLAASRPSWGSRTKCEDNQFQCRNGRCITSVWQCDGDDDCSDGTDELS
 CVKKTCAESDFVCSNGQCVPERWQCDGDQDCEDGSDEHPEFCHLRTRVTEISCGPRSTQCIPVSW
 SCDGESDCDSGEDETNCGNITCSPDEFTCSSGRCVSKIFFCNGHDDCGDGSDELDCPTPTCGAHEF
 QCKNSSCIPI SWVCDDADCADHSDESMEQCGRQPASPQRCPESEIQCGSGECIHKRWRCGDADC
 KDKSDEINCPPrTCRPDQFKCQDGNVCVHGSRQCNGVRDCLDGTDEINCKTVNQCSGPGKFKCRSGE
 CIEISKVCNELQDCRDWSDEPLKECGLNECLVNNGGCSHICRDLIVGYECECPAGFELIDRKTCDG
 IDECHNPGICSQICVNLKGGYKCECSRGYQMDLATGVCKAVGTEPCLIFTNRRDIRKIGLERKEYK
 QLVEQLRNTVALDADIAGQNI FWADVNQKSIFRASVDSRDKIGRHVKI INNLHSPASIAVDWVYKN
 IYWADTGLKTI SVSNFNFSRTKILIDTDLKEPTSI AVDPISGFIYWSWGEPARIEKAGMNGIDRQ
 KLVTTDIQWPNGIALDLVKSRLYWLDSKHLMLSSVDFNGQDRRIVLKSLEFLAHPLALTI FEDRVF
 WIDGENDAIYGANKFTGQEIATLVNNLIDAQDI IVYHELVQPSGKNWCNEGQDNGGCEYLCLPAPQ
 INSHSPKYTCACPAGSELEDGGGRCTKDVGGTAAYGETKDASTRGKSPIEEHSSSVFEANQSTKGS
 AAWAILPLLLLAMAAGGGYLVWRNWKRNKSMNFDNPVYLKTTEEDLNIDLGRHSTSVGHTYPAI
 SIVSTEDDLA

4.4 Analisi dell'espressione genica

Il numero di reads identificate durante la fase del mapping è stato fondamentale per effettuare l'analisi dell'espressione genica nei vari individui e tessuti (gonadi di entrambi i sessi e fegato femminile).

Individuo-tessuto	F1F	F2F	F3F	M1G	M2G	M3G	F1G	F2G	F3G
Reads	95	662	270	288	371	612	33016	31346	42878

Prima di procedere con l'analisi dell'espressione genica, è stato fondamentale confrontare i livelli delle reads ottenute con la concentrazione iniziale di RNA di partenza, per essere certi che la differenza nel numero di reads non fosse correlata alla differenza di quantità dell'RNA di partenza.

Individuo-tessuto	F1F	F2F	F3F	M1G	M2G	M3G	F1G	F2G	F3G
RNA (ng/μl)	455,0	570,9	483,3	86,0	91,0	64,0	251,6	424,0	813,5

Dal numero di reads identificate con *RNA-Seq mapping tool* nel CLC Genomics Workbench, è stato possibile identificare il valore di espressione di Seq1 nei singoli tessuti e individui, calcolando prima gli RPKM e successivamente i TPM.

Individuo-tessuto	F1G	F2F	F3F	M1G	M2G	M3G	F1G	F2G	F3G
RPKM	12,65	78,53	33,30	38,45	35,83	65,75	2642,27	2212,91	3149,27
TPM	7,78	49,68	22,33	18,92	17,84	33,19	983,89	827,86	1185,28

I valori di TPM sono poi stati utilizzati per effettuare una rappresentazione

grafica:

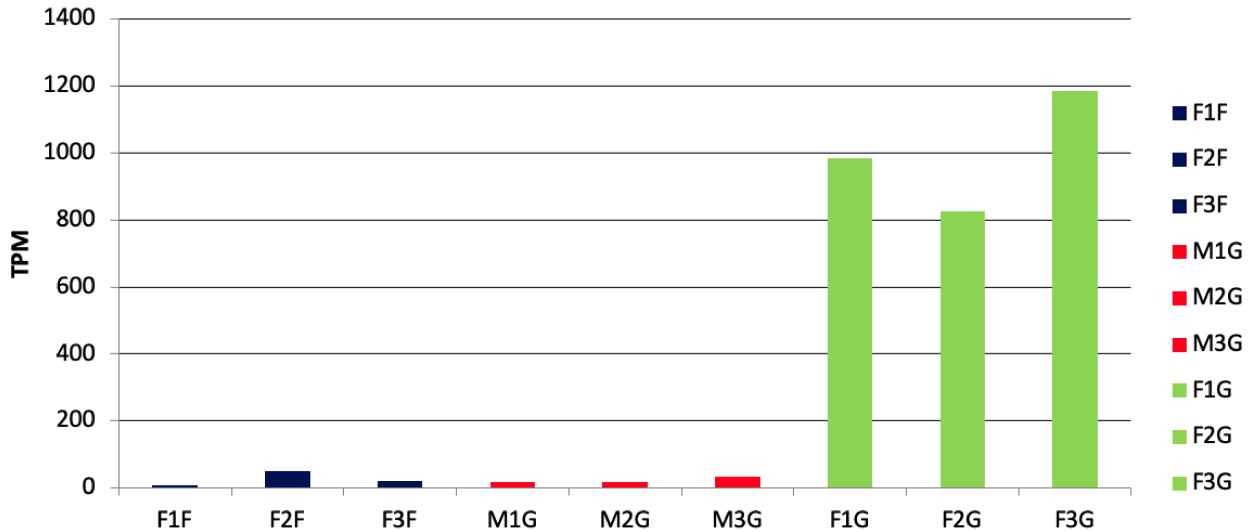


Figura 28. Grafico che mostra i livelli di espressione di Seq1 in termini di TPM: ogni colonna rappresenta un individuo; in verde le gonadi femminili, in rosso le gonadi maschili e in blu i fegati femminili.

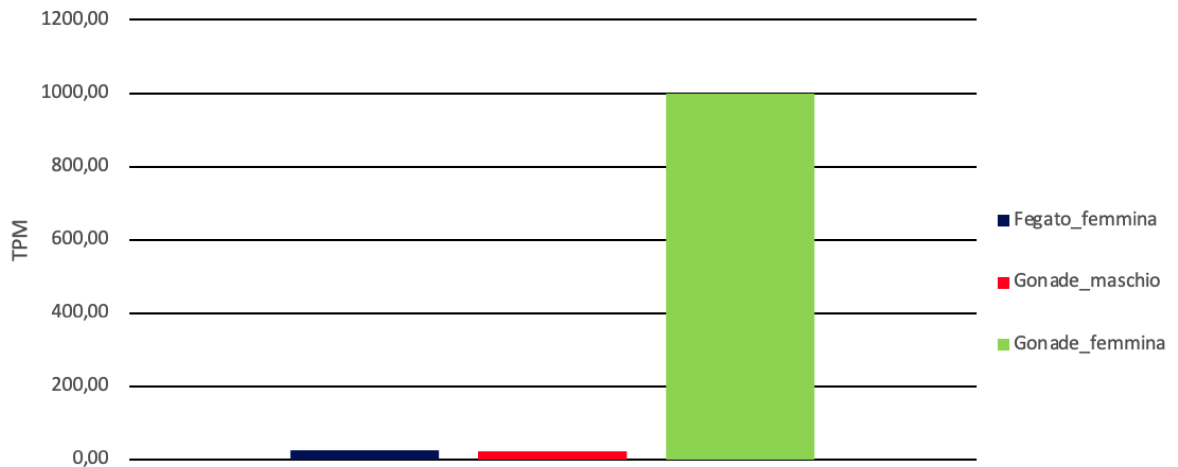


Figura 29. I livelli di espressione sono stati utilizzati per produrre una media considerando i tessuti: in verde la gonade femminile, in rosso la gonade maschile e in blu il fegato femminile.

L'analisi ha portato all'identificazione dei livelli di espressione di seq1 in

ogni individuo e tessuto: le gonadi femminili sono i tessuti che maggiormente

esprimono la sequenza presa in analisi; bassi livelli di espressione sono stati comunque identificati anche a livello delle gonadi maschili, paragonabili all'espressione avuta nel fegato femminile.

4.5 Analisi filogenetica

Ottenuta la sequenza completa di *seq1*, è stata effettuata un'analisi filogenetica utilizzando l'inferenza bayesiana mediante il programma MrBayes. Oltre alle sequenze ottenute da *C. orientalis* nell'analisi sono state inserite le sequenze della famiglia LdlR individuate da Reading e colleghi (2014). La ricerca dell'albero mediante MrBayes prende in considerazione un particolare valore, chiamato anche "temperatura dell'albero": è il valore di *Average standard deviation of split frequencies* (ASDSF) che deve raggiungere valori inferiori al 0,007 per permettere di identificare un albero attendibile dal punto di vista filogenetico. Poiché il valore di ASDSF non è sceso a sufficienza con 2,000,000 generazioni, è stato effettuato un ulteriore milione di generazioni, arrivando ad un valore di ASDSF di 0,004810 e un *Potential scale reduction factor* (PSRF) di 1,000, indicando appunto che il numero di generazioni totale di 3,000,000 era sufficiente. La ricerca dell'albero ha richiesto un tempo di 54 ore, 14 minuti e 36 secondi, ed ha portato i seguenti risultati:

- Seq1 ipotizzata come possibile sequenza del VtgR, è la più vicina evolutivamente al VtgR dello *X. laevis*, raggruppando anche con altre VldlR di *H. sapiens*, *M. musculus*, di *G. gallus* e la VtgR del *M. americana*.
- Seq2 raggruppa con tutte le LdlR prese in considerazione;
- Seq3 è la più simile alle Lrp8;
- Seq4 è evolutivamente più vicina alle Lrp2;
- Seq5 e Seq7 sono le uniche che si trovano in coppia a raggruppare con lo stesso tipo di proteine, le Lrp1;
- Seq6 raggruppa con le proteine identificate come Lrx+1.

Si nota quindi come ogni sequenza presa in considerazione va a raggruppare con proteine differenti (unica eccezione Seq5 e Seq7), mentre Seq1 è l'unica che risulta essere evolutivamente più vicino alle VtgR/VldlR (fig.30). I numeri sopra ad ogni nodo rappresentano la probabilità a posteriori: è la probabilità che l'albero ottenuto sia corretto considerando il modello evolutivistico scelto considerando la probabilità a priori (le informazioni già disponibili), realizzando quindi l'albero più probabile. Il fatto che la maggior parte dei nodi presentino una probabilità uguale alla certezza dell'evento (=1) rende attendibile le relazioni evolutive identificate nell'analisi.

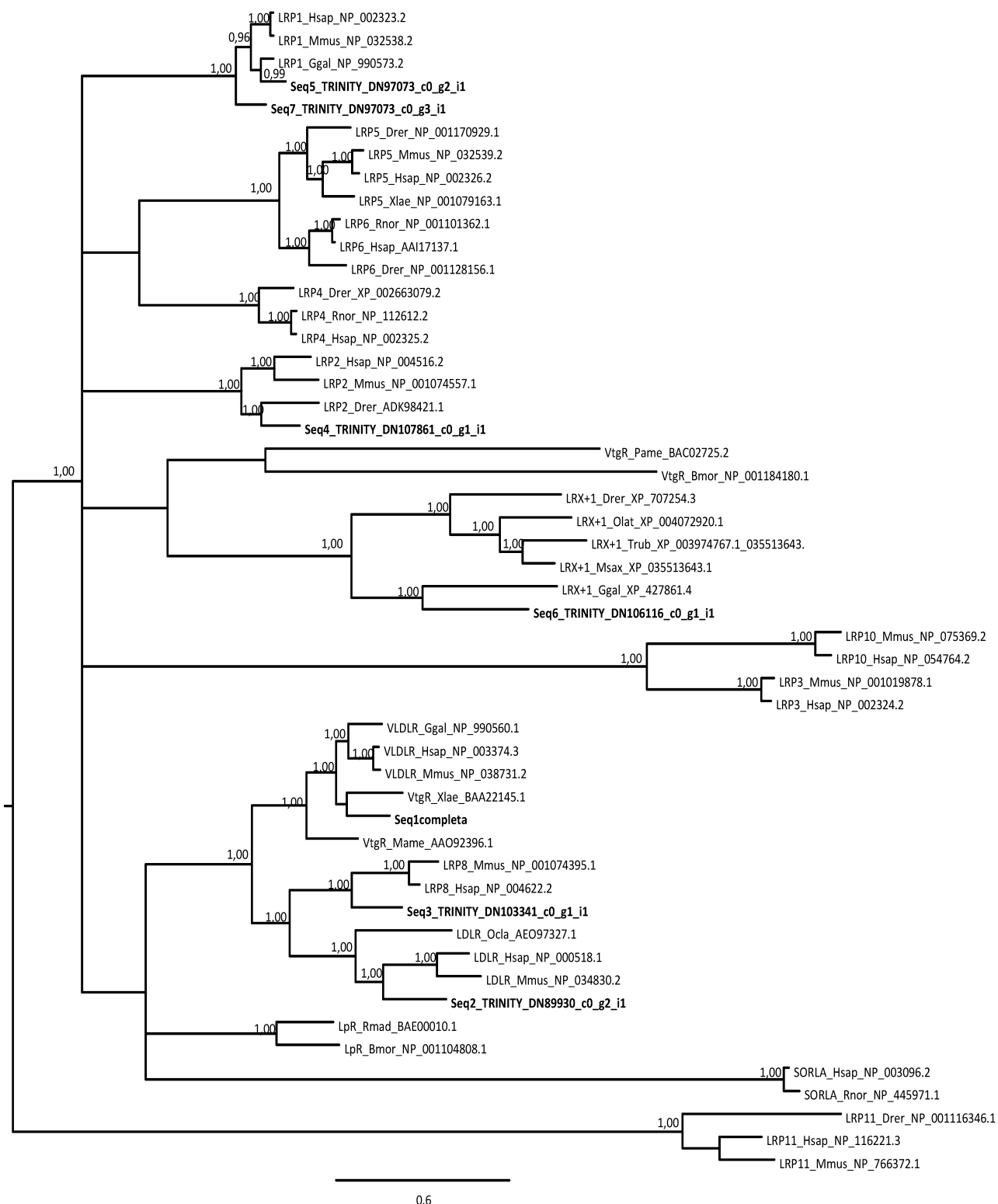


Figura 30. Albero filogenetico ottenuto con le sequenze prese in analisi da Reading e colleghi (2014) e dal database del trascrittoma: in grassetto sono evidenziate le sequenze della *C. orientalis*.

4.6 Predizione della struttura terziaria

Una volta ottenuta la sequenza amminoacidica tramite il *Translate tool* – *ExPASy*, si è effettuata una predizione di struttura terziaria andando ad utilizzare il server online di I-TASSER.

Come primo risultato il server riporta la propensione dei singoli amminoacidi nel realizzare strutture secondarie: la maggior parte della sequenza si trova organizzata in coil e non presenta quindi una particolare organizzazione strutturale; le uniche strutture secondarie riscontrate sono degli β -strands, concentrati per la maggior parte nella seconda metà della sequenza proteica.

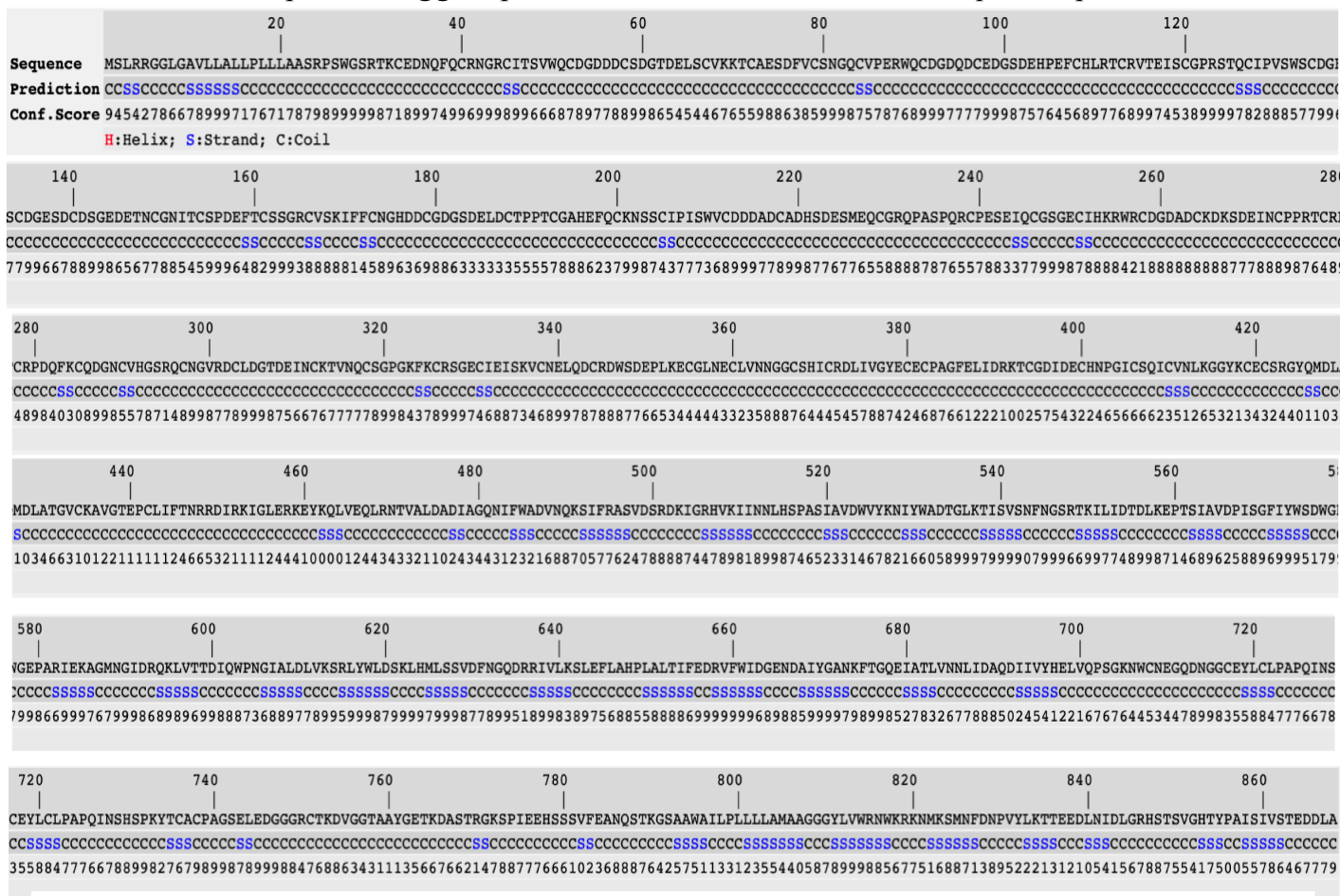


Figura 31. Predizione di struttura secondaria identificata da I-TASSER; la lettera C indica la propensione dell'amminoacido nel trovarsi in un coil; la lettera S indica la propensione dell'amminoacido di trovarsi in uno strand. Il numero indica la confidenza della predizione: più è alto il punteggio più è alta la confidenza della struttura secondaria.

La confidenza della predizione è rappresentata dal numero sotto ogni amminoacido e può variare da 0 a 9, dove 9 è la massima confidenza nel risultato. Il punteggio risulta essere adeguato nella maggior parte della proteina, eccetto in alcune porzioni come quella tra i 400 e i 480 amminoacidi dove non si può identificare con certezza assoluta una struttura secondaria. La rappresentazione grafica riporta la stabilità delle strutture secondarie secondo la normalizzazione del fattore B, un fattore che valuta la mobilità termica intrinseca nei residui delle proteine: si identificano delle porzioni più stabili e che corrispondono alle strutture β -strands.

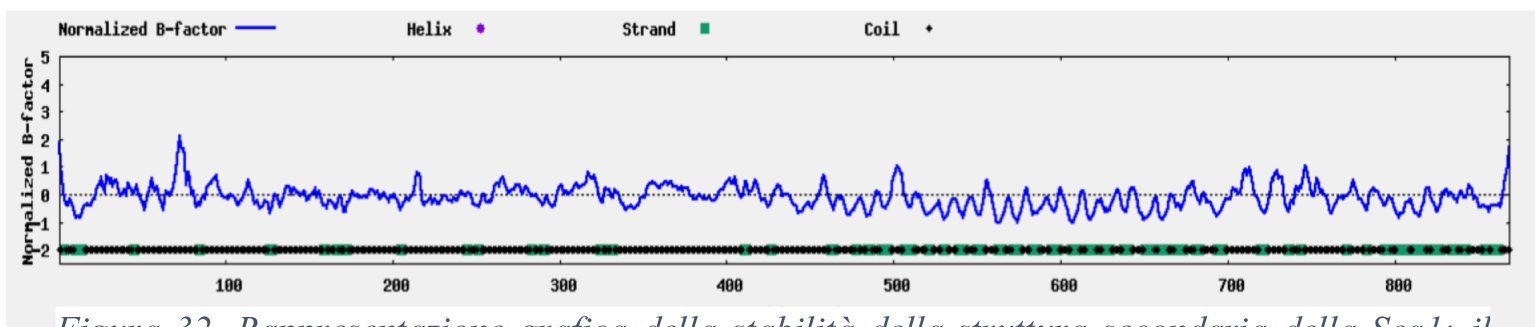


Figura 32. Rappresentazione grafica della stabilità della struttura secondaria della Seq1; il grafico è realizzato secondo la normalizzazione del fattore B identificato per ogni amminoacido: più aumenta, più il residuo è caratterizzato da una mobilità termica. In verde sono rappresentate le porzioni in strand, in nero quelle in coil.

Per la predizione della struttura terziaria, il server realizza cinque ipotetici modelli che potrebbero corrispondere al corretto ripiegamento di Seq1. Ciascun modello ha una confidenza espressa quantitativamente dal C-score identificato nelle fasi di ricerca nei database, ed ha un valore generalmente compreso tra -5 e 2: più il punteggio è alto, più si ha una maggiore confidenza nel modello. I

modelli proposti per Seq1 hanno tutti un C-score negativo, ma il modello con la confidenza più alta risulta essere il primo proposto.

Modello	1	2	3	4	5
C-score	-0,90	-2,92	-3,03	-2,07	-3,22

Analizzando il modello che presenta la miglior confidenza, si osserva come le porzioni in β -strand vanno ad organizzarsi in cinque β -sheets antiparalleli, composti ognuno da quattro β -strands, concentrati tutti in una porzione centrale e circondati dalle porzioni coils. In particolare, ogni β -sheet è disposto in modo molto regolare formando una struttura ciclica: solo quattro β -strands non sono organizzati come le altre porzioni, occupando solo parzialmente una posizione nella struttura ciclica.

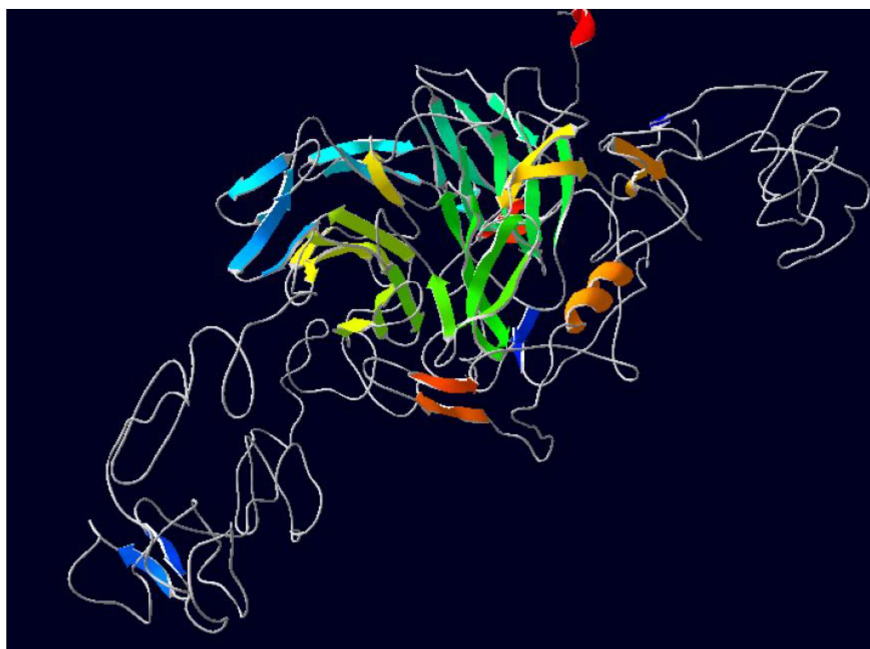


Figura 33. Struttura tridimensionale della Seq1 proposta dal modello1 analizzata con SwissPDBviewer (Guex et al., 1997): in grigio vengono mostrate le porzioni relative ai coils; dal blu al rosso, in successione dall' N-terminale al C-terminale, le porzioni organizzate in strutture secondarie ben definite.

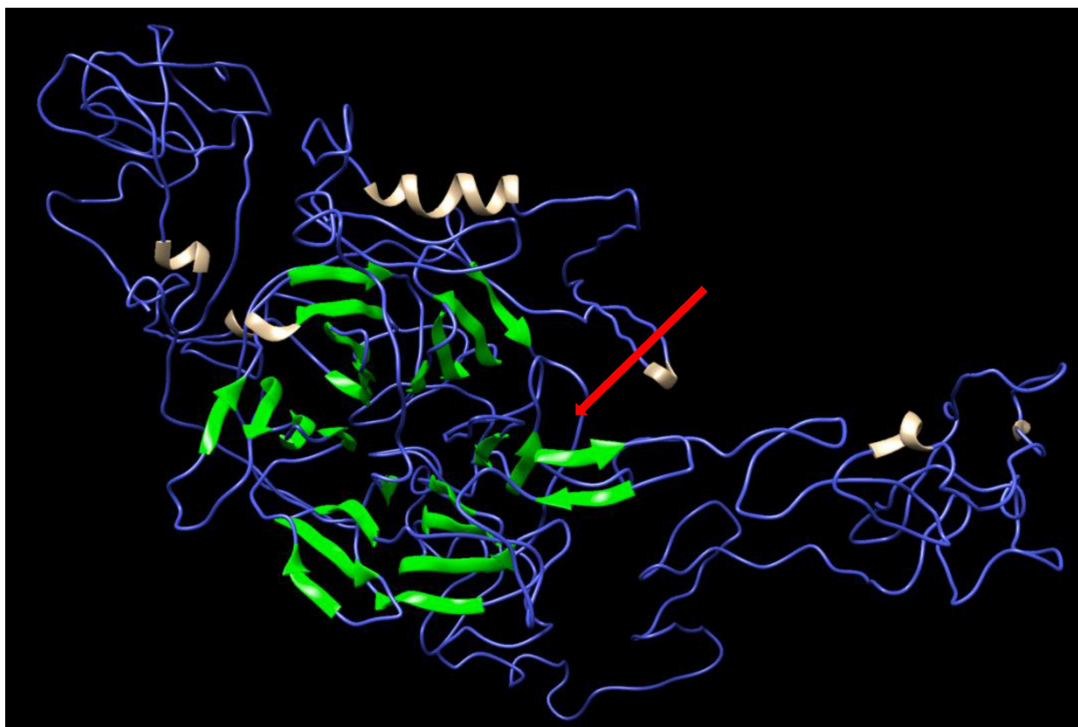


Figura 34. Rappresentazione del modello1 analizzata con Chimera (Pettersen et al.,2004): in verde la disposizione dei β -strands con una disposizione ciclica, in blu le porzioni relative ai coils, in ocra le alfa eliche; la freccia rossa indica le due coppie di β -strands non organizzate nel β -sheet.

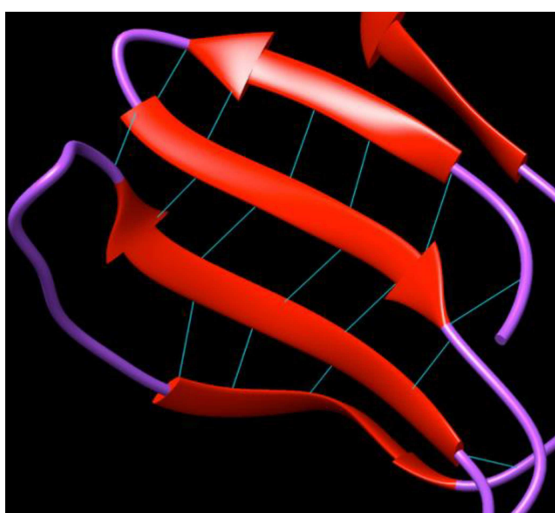


Figura 35. Rappresentazione di uno dei β -strands organizzati nelle porzioni centrali: in azzurro sono rappresentati i legami idrogeno tra le differenti porzioni.

Nelle analisi, il programma riconosce dieci proteine nel database del PDB che presentano una struttura paragonabile a quella di Seq1 identificata dal modello1. Successivamente le dieci proteine vengono organizzate in una classifica in base al TM-score che indica la somiglianza topologica delle due strutture proteiche: su dieci proposte, le proteine che appartengono alla famiglia delle LdlR sono sette, in particolare i modelli posti in posizione 1,2,3,4,6,8 e 10.

Top 10 Identified structural analogs in PDB

	Click to view	Rank	PDB Hit	TM-score	RMSD ^a	IDEN ^a	Cov	Alignment
LDLR	<input checked="" type="radio"/>	1	1n7dA	0.716	0.93	0.484	0.721	Download
LDLR	<input checked="" type="radio"/>	2	5b4xB	0.465	3.76	0.453	0.511	Download
LRP4	<input checked="" type="radio"/>	3	3v64C	0.371	2.17	0.389	0.386	Download
LRP6	<input checked="" type="radio"/>	4	6l6rA	0.356	2.84	0.274	0.376	Download
	<input type="radio"/>	5	5c2vA	0.351	4.71	0.078	0.409	Download
LDLR	<input checked="" type="radio"/>	6	3m0cC	0.348	1.53	0.353	0.356	Download
	<input type="radio"/>	7	7abiE	0.345	7.19	0.042	0.478	Download
LRP6	<input checked="" type="radio"/>	8	3s8vA	0.345	2.98	0.242	0.367	Download
	<input type="radio"/>	9	6llIA	0.342	7.45	0.048	0.482	Download
LRP6	<input checked="" type="radio"/>	10	3sobB	0.338	1.76	0.377	0.348	Download

Figura 36. Classifica delle dieci proteine identificate nel PDB: il TM-score indica la corrispondenza topologica tra il modello1 e la proteina del PDB, ha un valore tra 0 e 1 dove 1 indica la completa corrispondenza delle due strutture; l'RMSD^a esprime la similarità strutturale, più è basso e più le proteine sono simili nell'organizzazione spaziale; IDEN^a è la percentuale della regione strutturalmente allineata e Cov (coverage) è la copertura dell'allineamento ed è calcolato con il numero di residui allineati sulla lunghezza della query (identificati dal TM-align).



Figura 37. Rappresentazione del dominio extracellulare LdlR identificato nel PDB come il più simile al modello1 proposto: in arancione i β -strands organizzati in β -sheets; in azzurro le porzioni in coil.

5. Discussione

5. Discussione

La vitellogenina è un'apolipoproteina prodotta nel fegato femminile di organismi ovipari e ovovivipari (Verderame et al., 2017; Biscotti et al., 2018). La sua funzione principale è quella di prendere parte alla formazione del vitello nelle uova, rappresentando una riserva nutrizionale per lo sviluppo embrionale e larvale. In realtà si parla di famiglia genica di Vtg, derivate da duplicazioni di un unico gene ancestrale che hanno permesso la formazione di differenti Vtg. Il loro recettore, appartenente alla famiglia delle LdlR (Bujo et al., 1994; Davail et al., 1998), è fondamentale per l'endocitosi di Vtg all'interno degli oociti (Opresko et al., 1987): in alcuni uccelli a livello di questo recettore si sono osservate mutazioni o ablazioni genetiche capaci di portare l'ovario a uno sviluppo anormale, causandone l'infertilità (Bujo et al., 1995). Quindi espandere le conoscenze che si hanno delle Vtg e dei loro recettori permette non solo di comprendere a fondo il meccanismo della vitellogenesi ma anche di individuare nuove funzioni e processi metabolici in cui queste proteine sono coinvolte. In particolare, in questa tesi ci si è posti l'obiettivo di individuare e caratterizzare il recettore della vitellogenina in *Cynops orientalis*, specie di anfibio appartenente all'ordine Caudata. Si è scelto questo organismo perché negli anfibi sono stati identificati i geni *vtgR* solo all'interno degli Anuri, mentre nessuna sequenza è stata annotata e riportata nelle banche dati per

l'ordine dei Caudata. La *C. orientalis* appartiene agli organismi con il genoma più grande tra i vertebrati, fatta eccezione dei pesci polmonati (Sun et al., 2012), ed ha un genoma di circa 44,22Gb che rende difficoltosa la ricostruzione genomica e cromosomica, ad oggi non ancora disponibile. Per *C. orientalis*, allo stato attuale, è disponibile solo il trascrittoma (Biscotti et al., 2020) che ha rappresentato il punto di partenza di questa tesi. Dal trascrittoma è stato quindi possibile recuperare sequenze corrispondenti alle LdlR, definirne le relazioni filogenetiche, individuare la sequenza ortologa ai geni *vtgR* e su questa effettuare delle analisi più approfondite come la predizione di struttura terziaria e analisi quantitative di espressione.

Poiché *vtgR* appartiene a una famiglia genica di recettori LdlR (Bujo et al., 1994; Davail et al., 1998), è stato utile basare le analisi filogenetiche sullo studio realizzato precedentemente da Reading e colleghi (2014) che per effettuare la loro analisi filogenetica hanno preso in considerazione differenti sequenze LdlR per accertarsi che il nuovo recettore della vitellogenina identificato, Lrp13, non fosse un'isoforma di altre LdlR già individuate. Nell'albero filogenetico ottenuto in questa tesi si evidenzia la correlazione tra la Seq1 e le VtgR/VldlR di anfibi (*X. laevis*), pesci (*M. americana*), mammiferi (*H. sapiens* e *M. musculus*) e uccelli (*G. gallus*): le relazioni di ortologia sono rappresentate dalla topologia dei rami, dove, seguendo le relazioni

tassonomiche, VtgR dei pesci risulta avere una posizione più esterna rispetto ad anfibi, uccelli e mammiferi che si raggruppano separatamente. È possibile quindi confermare che Seq1 di *C. orientalis* è una VtgR ortologa alle altre prese in considerazione nell'analisi. Seq1 è l'unica sequenza tra quelle prese in considerazione dal trascrittoma che si colloca in questa posizione: si può dedurre quindi che *vtgR* in *C. orientalis* non si sia duplicato e che quindi è da escludere la presenza di più VtgR che possano interagire con le differenti Vtg identificate da Carducci e colleghi (2021) (*VtgCol*, *VtgCoII*, *VtgCoIII* e *VtgCoI*).

La correlazione tra Seq1 e le proteine LdlR si evidenzia anche attraverso i modelli di struttura terziaria predetti attraverso il server online I-TASSER: il modello1 proposto risulta avere una struttura molto simile a quella della proteina identificata in PDB come il dominio extracellulare LdlR nell' *H. sapiens*, classificandola quindi come proteina addetta al trasporto lipidico: c'è da considerare che all'interno di PDB non sono state inserite proteine VtgR, per cui è comprensibile individuare il dominio extracellulare della LdlR (PDB ID: 1N7D) come la struttura più simile a VtgR della *C. orientalis*. L'organizzazione delle due strutture risulta essere molto simile non solo per quanto riguarda la disposizione tridimensionale, bensì anche nelle strutture secondarie: le due proteine presentano un RMSD di 0,93Å ovvero una distanza

molto piccola per trattarsi di un modello predetto. Ciò ha permesso quindi di identificare il modello1 come un modello di predizione attendibile per VtgR della *C. orientalis*.

La posizione nell'albero filogenetico delle sequenze Seq1, 2 e 3 con le VldLR, Lrp8 e le sequenze indicate come LdlR, risulta essere particolare se si considerano i taxa analizzati: la presenza di tre geni nei pesci, uccelli, mammiferi e anfibi presi in considerazione fa ipotizzare che un unico gene ancestrale si sia duplicato in un antenato comune, prima della divisione tra Sarcopterigi e Attinopterigi. Questo gene potrebbe essere andato incontro ad una prima duplicazione formando *vtgR* e il gene progenitore per *LdlR/Lrp8*; quest'ultimo poi potrebbe essere stato soggetto ad una nuova duplicazione portando a due geni differenziati, Lrp8 e LdlR. Difatti, visto il coinvolgimento delle LdlR nel trasferimento dei lipidi all'interno degli oociti in maturazione nell' *O. clarki* (Luo et al., 2013), delle Lrp8 nelle gonadi per la riserva di colesterolo utile alla realizzazione di steroidi anche nella follicologenesi (Argov et al., 2004) e nel raggiungimento del giusto peso del tuorlo e albume (Wang et al., 2013), è possibile ipotizzare anche una loro interazione con Vtg, dato che anche altri recettori appartenenti a questa famiglia genica, come Lrp13, hanno la possibilità di legare le Vtg (Reading et al., 2014).

Le altre sequenze del trascrittoma prese in considerazione hanno mostrato correlazione con sequenze di altre LdlR: Seq4 corrisponde a Lrp2 identificata nei pesci e nei mammiferi; Seq6 è stata identificata come Lrx+1 mentre Seq5 e Seq7 raggruppano con le Lrp1, evidenziando in *C. orientalis* una possibile duplicazione di questo gene. Tuttavia, non essendoci un gruppo costituito solo da Seq5 e Seq7, le due sequenze non sono le più simili tra loro e quindi la duplicazione non è esclusiva degli anfibi ma potrebbe essere avvenuta, anche in questo caso, in un progenitore ancestrale tra uccelli (*G. gallus*) mammiferi (*H. sapiens* e *M. musculus*) e anfibi (*C. orientalis*). Si evidenzia quindi la necessità di uno studio più approfondito per la classificazione di queste due sequenze. Nessuna sequenza nel trascrittoma della *C. orientalis* ha raggruppato con Lrp3, Lrp4, Lrp6, Lrp5, Lrp10 o Lrp11: la spiegazione della loro assenza è probabilmente legata al fatto che i tessuti considerati per la realizzazione del trascrittoma (fegato e gonadi) potrebbero non esprimere o esprimere a bassi livelli questi geni e i trascritti, oppure ad una loro perdita nelle fasi di retrotrascrizione e assemblaggio *de novo*. Non è opportuno credere che questi geni siano assenti nella *C. orientalis* poiché mancano studi approfonditi (anche genomici) che abbiano identificato od escluso la presenza di questi recettori negli anfibi.

Parallelamente, è stato valutato il livello di espressione del gene *vtgR* della *C. orientalis*. Considerando i valori degli RPKM ottenuti per *vtgR* è stato possibile confermare la sua espressione limitata principalmente nella gonade femminile, anche se bassi livelli sono stati identificati nel fegato femminile e nella gonade maschile. Poiché Vtg è coinvolta anche in altri processi differenti dalla vitellogenesi (Sun et al., 2015) e può essere espressa anche a livello di individui maschili (Verderame et al., 2014), l'espressione di VtgR a livello della gonade maschile potrebbe rappresentare un coinvolgimento del recettore anche in funzioni non correlate alla vitellogenesi. In *G. gallus* l'espressione di VldLR/VtgR è stata identificata, a bassi livelli, anche in altri tessuti differenti dalla gonade, ovvero nel cuore e nel muscolo scheletrico (Bujo et al., 1994), tessuti che nei mammiferi più esprimono le VldLR (Takahashi et al., 1992). In questa tesi è stato analizzato un numero limitato di tessuti e quindi non è da escludere l'espressione di *vtgR* anche in altri tessuti.

I livelli di espressione dei geni *vtg* identificate nella *C. orientalis* (Carducci et al, 2021) sono stati confrontati con i livelli di espressione di *vtgR* identificati in questa tesi e con le sezioni istologiche realizzate sulle stesse gonadi femminili da Biscotti e colleghi (2020). Da tale comparazione è emersa una correlazione tra i livelli di espressione dei geni e lo stadio di maturazione gonadico. Poiché nella *C. orientalis* sono state identificate due vitellogenine, VtgCoII e

VtgCoIV, come possibili responsabili della formazione della riserva nutrizionale negli oociti (Carducci et al, 2021), è stato possibile ipotizzare una correlazione tra i livelli di espressione dei geni *vtgCoII* e *vtgCoIV* e *vtgR*. La gonade della femmina 3 (F3G) è quella che più esprime *vtgR* (1,185.28 TPM) ma risulta anche essere l'individuo con una gonade immatura e con bassi livelli di espressione dei geni *vtg*, poiché è stato osservato che VtgR ha un'espressione pre-vitellogenica (Perazzolo et al., 1999). E' possibile quindi attribuire uno stadio pre-vitellogenico agli oociti presenti nella gonade della F3, che quindi esprimono *vtgR* per prepararsi alla produzione del vitello, osservazione anche confermata dall'analisi delle sezioni gonadiche. La gonade della femmina 2 (F2G), essendo caratterizzata da oociti più maturi, ha un basso livello di espressione dei geni *vtg* e una minor espressione del recettore (827.86 TPM) poiché ha già accumulato vitello negli oociti. La gonade della femmina 1 (F1G) è invece in uno stadio intermedio, in cui presenta sia oociti vitellogenici che pre-vitellogenici, con un'alta espressione dei geni *vtg* e un livello di espressione per il recettore (983.89 TPM) intermedio tra F3G e F2G.

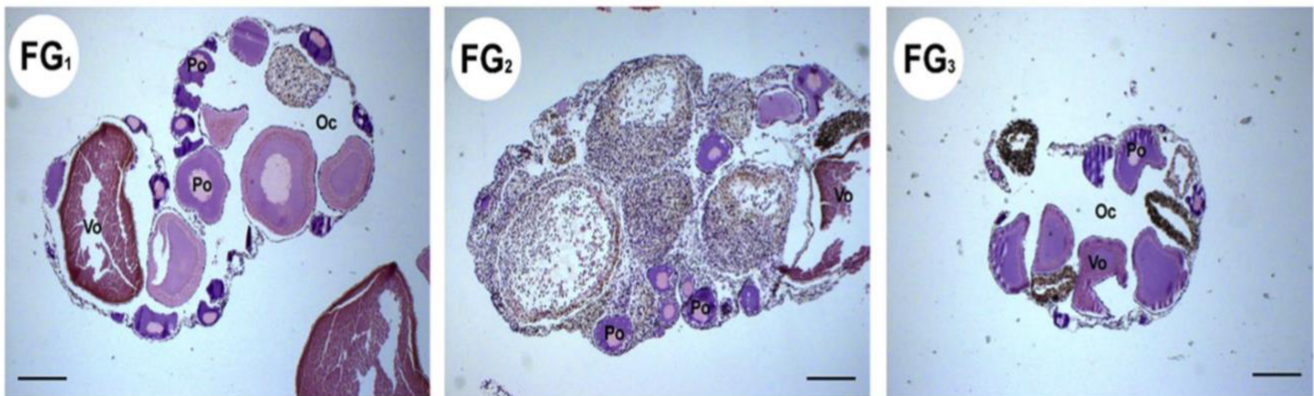


Figura 38. Sezione istologica delle gonadi di due individui di *C. orientalis* (Biscotti et al., 2020): le gonadi femminili con oociti a differente stadio di maturazione racchiusi nella cavità ovarica (Oc), oocita vitellogenico (Vo) e oocita previtellogenico (Po); la barra nera rappresenta la scala da 400 μ m.

In conclusione, si è potuta determinare la corrispondenza di sequenze del trascrittoma di *C. orientalis* con recettori appartenenti alla famiglia delle LdlR. La realizzazione dell'albero filogenetico ha chiarito le relazioni di ortologia tra le sequenze, permettendoci di identificare la sequenza *vtgR* in *C. orientalis*. Inoltre, i risultati suggeriscono l'ipotesi che *vtgR/vldlR* e *lrp8* siano derivati da duplicazioni a partire da un unico gene nell'antenato comune di Attinopterigi e Sarcopterigi. L'analisi della struttura terziaria di Seq1/VtgR di *C. orientalis* ha permesso di identificare una corrispondenza con i domini LdlR registrati all'interno del PDB. Questo studio può rappresentare quindi un punto di partenza per successivi approfondimenti sulle funzioni di VtgR, degli altri recettori appartenenti alle LdlR e dei diversi geni della famiglia delle Vtg negli anfi.

6. Ringraziamenti

Al termine di questo percorso, sento di dover ringraziare molte persone che mi sono state vicine e che mi hanno sostenuta passo dopo passo in questo cammino.

Vorrei ringraziare il mio relatore, il professor Marco Barucca, per avermi guidata nella ricerca e nella realizzazione di questo elaborato.

Un grazie speciale a Fede, la mia correlatrice, che con dolcezza e molta pazienza mi ha sempre affiancato, pronta a spiegarmi (e rispiegarmi) ogni passaggio, dimostrandosi una persona speciale e di grande professionalità.

Grazie alle professoresse Maria Assunta Biscotti e Adriana Canapa per i suggerimenti pratici che mi hanno dato nell'intero percorso e ad Elisa Carotti che è sempre stata disponibile per qualsiasi mio dubbio.

Vorrei ringraziare di cuore i miei genitori per avermi aiutata a crescere e a diventare la persona che sono oggi: grazie a mia madre che è sempre stata al mio fianco in tutte le sfide che ho affrontato, mio unico esempio di coraggio e perseveranza, e a mio padre che mi ha insegnato il bene incondizionato.

Un grazie ai miei fratelli, Federico e Simone, a cui voglio un mondo di bene, per essere sempre pronti a tirarmi su di morale.

Grazie a Samuele che, sottovoce, mi ha insegnato il valore della pazienza e del rispetto e che riesce sempre a tirare fuori il meglio di me.



7. Bibliografia

- Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J.: Basic local alignment search tool. *Journal of molecular biology* (1990);215(3):403-410;
- Andruch V., Kocúrová L., Balogh S.I., Škrliková J.: Recent advances in coupling single-drop and dispersive liquid–liquid microextraction with UV–vis spectrophotometry and related detection techniques. *Microchemical Journal* (2012); 102:1-10;
- Argov N., Moallem U., Sklan D.: Lipid transport in the developing bovine follicle: messenger RNA expression increases for selective uptake receptors and decreases for endocytosis receptors. *Biol Reprod.* (2004);71(2):479-485;
- Barucca M., Biscotti M.A., Forconi M., Regoli F., Olmo E., Canapa A.: Characterization and phylogenetic analysis of vitellogenin coding sequences in the Antarctic fish *Trematomus bernacchii*. *Journal of experimental zoology. Part B, Molecular and developmental evolution* (2010);314(8): 645-52;
- Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E.: The Protein Data Bank. *Nucleic Acids Res* (2000); 28(1):235-42;

- Bhandari R.K., Taniyama S., Kitahashi T., Ando H., Yamauchi K., Zohar Y., Ueda H., Urano A.: Seasonal changes of responses to gonadotropin-releasing hormone analog in expression of growth hormone/prolactin/somatolactin genes in the pituitary of masu salmon. *General and comparative endocrinology* (2003);130(1):55-63;
- Biscotti M. A., Barucca, M., Carducci, F., Canapa, A.: New perspectives on the evolutionary history of vitellogenin gene family in vertebrates. *Genome Biology and Evolution*(2018)., 10: 2709–2715;
- Biscotti M.A., Carducci F., Barucca M., Gerdol M., Pallavicini A., Scharl M., Canapa A, Adolfi M.C.: The transcriptome of the newt *Cynops orientalis* provides new insights into evolution and function of sexual gene networks in sarcopterygians. *Sci Rep* (2020);10:5445;
- Blaunstein A.R., Kiesecker, J.M.: Complexity in conservation: lessons from the global decline of amphibian population. *Ecology Letters* (2002); 5:597-608;
- Bolger A.M., Lohse M., Usadel B.: Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* (2014); 30(15):2114-20;
- Bujo H., Hermann M., Kaderli M.O., Jacobsen L., Sugawara S., Nimpf J., Yamamoto T., Schneider W. J.: Chicken oocyte growth is mediated

- by an eight ligand binding repeat member of the LDL receptor family. *The EMBO journal* (1994);13(21): 5165-75;
- Bujo, H., Yamamoto, T., Hayashi, K., Hermann, M., Nimpf, J., Schneider, W. J.: Mutant oocytic low density lipoprotein receptor gene family member causes atherosclerosis and female sterility. *Proc. Natl. Acad. Sci.* (1995);92:9905–9909;
 - Canapa A., Olmo E., Forconi M., Pallavicini A., Makapedua M.D., Biscotti M.A., Barucca M.: Composition and phylogenetic analysis of vitellogenin coding sequences in the Indonesian coelacanth *Latimeria menadoensis*. *Journal of experimental zoology. Part B, Molecular and developmental evolution* (2012); 318(5): 404-16;
 - Carducci F., Biscotti M.A., Canapa A., Barucca M.: The vitellogenin genes in *Cynops orientalis*: New insights on the evolution of the vtg gene family in amphibians. *Journal of experimental zoology. Part B, Molecular and developmental evolution* (2021); 336(7): 554-561;
 - Carducci F., M. A. Biscotti & A. Canapa: Vitellogenin gene family in vertebrates: evolution and functions. *The European Zoological Journal* (2019); 86(1):233-240;
 - Chi W., Gan X., Xiao W., Wang W., He S.: Different evolutionary patterns of hypoxia-inducible factor α (HIF- α) isoforms in the basal

- branches of Actinopterygii and Sarcopterygii. *FEBS Open Bio.* (2013);3:479-83;
- Davail B., Pakdel F., Bujo H., Perazzolo L.M., Waclawek M., Schneider W.J., Le Menn F.: Evolution of oogenesis: the receptor for vitellogenin from the rainbow trout. *Journal of Lipid Research* (1998) 39:1929-1937;
 - Finn R.N., Kolarevic J., Kongshaug H., Nilsen F.: Evolution and differential expression of a vertebrate vitellogenin gene cluster. *BMC Evol Biol* (2009). 9(2);
 - Finn R.N., Kristoffersen B.A.: Vertebrate vitellogenin gene duplication in relation to the "3R hypothesis": correlation to the pelagic egg and the oceanic radiation of teleosts. *PloS one* (2007);2(1): e169;
 - Frohman M.A., Dush M.K., Martin G.R.: Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proceedings of the National Academy of Sciences of the United States of America* (1988); 85(23):8998-9002;
 - Frost et al.: The amphibian tree of life. *Bulletin of the American Museum of Natural History* (2006) ; 297:1-291;
 - Gao K.Q., Shubin N.H.: Late Jurassic salamanders from northern China. *Nature* (2001); 6828:574-7;

- Goldstein J.L., Brown M.S., Anderson R.G., Russell D.W., Schneider W.J.: Receptor-mediated endocytosis: concepts emerging from the LDL receptor system. *Annu Rev Cell Biol* (1985); 1:1–39;
- Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A.: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* (2011);29(7):644-652;
- Guex N., Peitsch M.C.: SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* (1997);18 (15):2714-2723;
- Hara A., Hiramatsu N., Fujita T.: Vitellogenesis and choriogenesis in fishes. *Fish Sci* (2016); 82(2): 187-202;
- Herz J: Deconstructing the LDL receptor – a rhapsody in pieces *Nature Structural Biology*(2001); 8: 476-8;
- Hiramatsu N., Todo T., Sullivan C.V., Schilling J., Reading B.J., Matsubara T., Ryu Y.W., Mizuta H., Luo W., Nishimiya O., Wu M., Mushirobira Y., Yilmaz O., Hara A.: Ovarian yolk formation in fishes:

- molecular mechanisms underlying formation of lipid droplets and vitellogenin-derived yolk proteins. *Gen. Comp. Endocrinol* (2015). 221: 9–15;
- Huelsenbeck J.P., Ronquist F.: MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* (2001); 17(8):754-755;
 - Jorgensen P.: Yolk. *Current biology* (2008);18(3): R103-4;
 - Katoh K. , Standley D.M.: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* (2013);30(4) :772-780;
 - Koduri V., Blacklow S.C.: Folding Determinants of LDL Receptor Type A Modules. *Biochemistry* (2001); 40,(43):12801–12807;
 - Lassmann T., Sonnhammer E.L.: Kalign, Kalignvu and Mumsa: web servers for multiple sequence alignment. *Nucleic Acids Research* (2006); 34(2): W596–W599;
 - Li A., Sadasivam M., Ding J.L. : Receptor-ligand interaction between vitellogenin receptor (VtgR) and vitellogenin (Vtg), implications on low density lipoprotein receptor and apolipoprotein B/E. The first three ligand-binding repeats of VtgR interact with the amino-terminal region of Vtg. *The Journal of biological chemistry* (2003); 278(5):2799-806;

- Li Y., Zhang Y.: REMO: A new protocol to refine full atomic protein models from C-alpha traces by optimizing hydrogen-bonding networks. *Proteins* (2009);76(3):665-76;
- Li Z., Zhang S., Liu Q. Vitellogenin functions as a multivalent pattern recognition receptor with an opsonic activity. *PLoS One*. (2008);3(4):e1940;
- Lu H.L., Wu Q., Geng, J., Dang W.: Swimming performance and thermal resistance of juvenile and adult newts acclimated to different temperatures. *Acta Herpetologica* (2016); 11(2):189-195;
- Luo W., Ito Y., Mizuta H., Massaki K., Hiramatsu N., Todo T., Reading B.J., Sullivan C.V., Hara A.: Molecular cloning and partial characterization of an ovarian receptor with seven ligand binding repeats, an orthologue of low-density lipoprotein receptor, in the cutthroat trout (*Oncorhynchus clarki*). *Comp Biochem Physiol A Mol Integr Physiol.*(2013);166(2):263-71;
- Madeira F., Park Y.M., Lee J., Buso N., Gur T., Madhusoodanan N., Basutkar P., Tivey A.R.N., Potter S.C., Finn R.D., Lopez R.: The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* (2019) 47(1): W636–W641;

- Metcalfe C.J., Casane D. : Accommodating the load: The transposable element content of very large genomes. *Mobile genetic elements*; (2013); 3,issue 2: e24775-1- e244775-8;
- Mizuta H., Luo W., Ito Y., Mushirobira,Y., Todo T., Hara A., Reading B.J., Sullivan C.V., Hiramatsu N.: Ovarian expression and localization of a vitellogenin receptor with eight ligand binding repeats in the cutthroat trout (*Oncorhynchus clarki*). *Comp. Biochem. Physiol. B* (2013);166 (1):81–90;
- Morini M., Lafont A. G., Maugars G., Baloché S., Dufour S., Asturiano J.F., Perez L. : Identification and stable expression of vitellogenin receptor through vitellogenesis in the European eel. *Animal* (2020); 14(6): 1213-1222;
- Mushirobira Y., Mizuta H., Luo W., Todo T., Hara A., Reading B.J., Sullivan C.V., Hiramatsu N.: Molecular cloning and partial characterization of a low-density lipoprotein receptor-related protein 13 (Lrp13) involved in vitellogenin uptake in the cutthroat trout (*Oncorhynchus clarki*). *Molecular reproduction and development*(2015); 82(12): 986-1000;
- Mushirobira Y., Niida M., Hotta T., Fujinami Y., Soyano K.: Hepatic expression profiles of three subtypes of vitellogenin and estrogen

- receptor during vitellogenesis in cultured female yellowtail. *General and comparative endocrinology* (2020); 299:113612;
- Notredame C., Higgins D.G., Heringa J.: T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* (2000); 302(1):205-17;
 - Nowakowski A.J., Watling J.I., Whitfield S.M., Todd B.D., Kurz D.J., Donnelly M.A.: Tropical amphibians in shifting thermal landscapes under land-use and climate change. *Conserv Biol* (2017);31:96-105;
 - Okabayashi K., Shoji H., Nakamura T., Hashimoto O., Asashima M., Sugino H.: cDNA cloning and expression of the *Xenopus laevis* vitellogenin receptor. *Biochem Biophys Res Commun.*(1996); 224(2):406-13;
 - Opresko L.K., Wiley H.S.: Receptor-mediated endocytosis in *Xenopus* oocytes. I. Characterization of the vitellogenin receptor system. *J Biol Chem* (1987) 262(9):4109-15;
 - Patiño R., Sullivan C.V. : Ovarian follicle growth, maturation, and ovulation in teleost fish. *Fish Physiology and Biochemistry*(2002). 26: 57–70;
 - Perazzolo L.M., Coward K., Davail B., Normand E., Tyler C.R., Pakdel F., Schneider W.J., Le Menn F.: Expression and localization of

- messenger ribonucleic acid for the vitellogenin receptor in ovarian follicles throughout oogenesis in the rainbow trout, *Oncorhynchus mykiss*. *Biology of reproduction* vol. 60,5 (1999):1057-68;
- Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E.: UCSF Chimera: a visualization system for exploratory research and analysis. *J Comput Chem.* (2004);25(13):1605-12;
 - Pough F. Harvey: Amphibian Biology and Husbandry. *ILAR Journal* (2007); 48 (3):203–213;
 - Pough, F. Harvey: The Advantages of Ectothermy for Tetrapods. *The American Naturalist* (1980); 1: 92–112;
 - Reading B.J., Hiramatsu N., Sawaguchi S., Matsubara T., Hara A., Lively M.O., Sullivan C.V.: Conserved and variant molecular and functional features of multiple egg yolk precursor proteins (vitellogenins) in white perch (*Morone americana*) and other teleosts. *Marine biotechnology* (2009);11(2): 169-87;
 - Reading B.J., Hiramatsu N., Schilling J., Molloy K.T., Glassbrook N., Mizuta H., Luo W., Baltzegar D.A., Williams V.N., Todo T., Hara A., Sullivan C.V.: Lrp13 is a novel vertebrate lipoprotein receptor that binds

- vitellogenins in teleost fishes. *Journal of lipid research* (2014); 55(11): 2287-95;
- Romano M., Limatola E.: Oocyte plasma membrane proteins and the appearance of vitellogenin binding protein during oocyte growth in the lizard *Podarcis sicula*. *General and comparative endocrinology* (2000); 118(3):383-92;
 - Saiki R.K., Gelfand D.H., Stoffel S., Scharf S.J., Higuchi R., Horn G.T., Mullis K.B., Erlich H.A.: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* (1988) 239(4839):487-91;
 - Shen Y., Chen Y.Z., Lou Y.H., Zhang C.X.: Vitellogenin and Vitellogenin-Like Genes in the Brown Planthopper. *Frontiers in physiology* (2019);10:1181;
 - Sievers F., Wilm A., Dineen D., Gibson T.J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J., Thompson J.D., Higgins D.G.: Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* (2011); 7:539;
 - Simão, F.A., Waterhouse R.M., Ioannidis P., Kriventseva E. V., Zdobnov, E.M.: BUSCO: assessing genome assembly and annotation

- completeness with single-copy orthologs. *Bioinforma. Oxf. Engl* (2015); 31: 3210–3212;
- Smith J.T.: Kisspeptin signalling in the brain: Steroid regulation in the rodent and ewe. *Brain reseach reviews* (2008); 57(2): 288-298;
 - Sparreboom M., Manuel Mouta Faria: Sexual behaviour of the Chinese fire-bellied newt, *Cynops orientalis*. *Amphibia-Reptilia* (1997); 18: 27-38;
 - Stecher G., Tamura K., Kumar S.: Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution* (2020);37(4):1237-1239;
 - Stifani S., Nimpf J., Schneider W.J.: Vitellogenesis in *Xenopus laevis* and chicken: cognate ligands and oocyte receptors. The binding site for vitellogenin is located on lipovitellin I. *Journal of biological chemistry* (1990); 265(2): 882-888;
 - Sun C, Zhang S.: Immune-Relevant and Antioxidant Activities of Vitellogenin and Yolk Proteins in Fish. *Nutrients* (2015);7(10):8818-29;
 - Sun C., López Arriaza J.R., Mueller R.L.: Slow DNA Loss in the Gigantic Genomes of Salamanders. *Genome Biology and Evolution* (2012);4(12):1340–1348;

- Takahashi S., Kawarabayasi Y., Nakai T., Sakai J., Yamamoto T:
Rabbit very low density lipoprotein receptor: a low density lipoprotein
receptor-like protein with distinct ligand specificity. *Proc. Natl. Acad.
Sci.* (1992);89: 9252-9256;
- Verderame M., Scudiero R. :Estrogen-dependent, extrahepatic synthesis
of vitellogenin in male vertebrates: A mini-review. *Comptes rendus
biologies* (2017);340(3): 139-144;
- Vieites D.R., P. Zhang, Wake D.B.: Salamanders (Caudata). *The
Timetree of Life* (2009); 365–368;
- Vitt Laurie and Janalee P. Caldwell: Herpetology: an introductory
biology of amphibians and reptiles. Academic press, 2014;
- Wahli W., Dawid I.B., Ryffel G.U., Weber R.: Vitellogenesis and the
vitellogenin gene family. *Science*(1981);212(4492): 298-304;
- Wake D. B.,Koo M. S.: Primer: *Amphibians*. *Current Biology* (2018);
28:1237-1241;
- Wallace R.A., Misulovin Z.: Long-term growth and differentiation of
Xenopus oocytes in a defined medium. *Proc.Nati.Acad.Sci.USA.*
(1978);75(11):5534-5538;
- Wang C., Li S.J., Li C., Yu G.H., Feng Y.P., Peng X.L., GongY.Z.:
Molecular cloning, expression and asso- ciation study with

- reproductive traits of the duck LRP8 gene. *British Poultry Science* (2013); 54: 567–574;
- Wu S, Zhang Y. : LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* (2007);35(10):3375-82;
 - Wu S., Zhang Y.: LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Research* (2007); 35: 3375-3382;
 - Yang J., Yan R., Roy A., Xu D., Poisson J., Zhang Y.: The I-TASSER Suite: Protein structure and function prediction. *Nature Methods* (2015); 12: 7-8;
 - Ye J., Coulouris G., Zaretskaya I., Cutcutache I., Rozen S., Madden T.L.: Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* (2012);13:134;
 - Zhang Y., Skolnick J.: SPICKER: A Clustering Approach to Identify Near-Native Protein Folds. *Journal of Computational Chemistry* (2004); 25: 865-871;
 - Zhang Y., Skolnick J.: TM-align: A protein structure alignment algorithm based on TM-score. *Nucleic Acids Research* (2005); 33: 2302-2309.