



UNIVERSITÀ POLITECNICA DELLE MARCHE

DIPARTIMENTO SCIENZE DELLA VITA E DELL'AMBIENTE

**Corso di Laurea Magistrale
BIOLOGIA MOLECOLARE E APPLICATA**

**ANALISI DEL DNA RIPETITIVO IN *CYNOPS ORIENTALIS*
ANALYSIS OF REPETITIVE DNA IN *CYNOPS ORIENTALIS***

Tesi di Laurea Magistrale di:
Dott.ssa Livianna Carrieri

Relatore:
Prof.ssa Maria Assunta Biscotti

Correlatore:
Dott.ssa Federica Carducci

**Sessione Estiva (Luglio)
Anno Accademico 2019-2020**

*“Nulla due volte accade
né accadrà. Per tal ragione
si nasce senza esperienza,
si muore senza assuefazione. [...]”*

-Wisława Szymborska

Sommario

Capitolo Primo.....	5
INTRODUZIONE	5
1.1. Scopo della tesi.....	5
1.2. Caudata	8
1.2.1. <i>Cynops orientalis</i>	12
1.2.2. Taglia e composizione genomica dei Caudata	14
1.2.2.1. DNA ripetitivo	16
1.2.2.2. Sequenze ripetute in tandem	17
1.2.2.3. Sequenze ripetute intersperse.....	20
Capitolo secondo.....	27
MATERIALI E METODI	27
2.1. Campionamento.....	27
2.2. Estrazione del DNA totale mediante alcool isoamilico e cloroformio (CIA)	27
2.3. Quantizzazione del DNA estratto con Qubit™	30
2.4. Costruzione della libreria genomica	31
2.4.1. Trattamento con RNasi.....	32
2.4.2. Digestione del DNA mediante endonucleasi di restrizione	32
2.4.3. Selezione dei frammenti di interesse da clonare e purificazione degli acidi nucleici da gel di agarosio	33
2.4.4. Clonaggio	34
2.4.5. Ligazione.....	37
2.4.6. Trasformazione mediante shock termico	38
2.4.7. Semina.....	38
2.5. Screening della libreria.....	39
2.5.1. Marcatura del DNA genomico di <i>C. orientalis</i>	39
2.5.2. Quantificazione della sonda e determinazione dell'efficienza della reazione di marcatura	40
2.5.3. Trasferimento delle colonie della libreria su membrana.....	41
2.5.4. Ibridazione del genoma marcato sulla membrana.....	42
2.5.5. Lavaggi stringenti post-ibridazione.....	43

2.5.6.	Determinazione immunologica	43
2.6.	Screening delle colonie di interesse mediante reazione a catena della polimerasi (PCR) 44	
2.7.	Estrazione plasmidica	47
2.8.	Sequenziamento	49
2.9.	Southern blotting	50
2.9.1.	Blotting.....	50
2.9.2.	Fissaggio del DNA alla membrana	51
2.9.3.	Marcatura con digossigenina DIG dell'inserito del clone 59p1	51
2.10.	Dot Blot quantitativo	52
2.11.	Analisi bioinformatica delle sequenze	54
2.11.1.	Dotlet.....	54
2.11.2.	Ricerca delle sequenze nelle banche dati	55
2.11.3.	Analisi delle sequenze con il programma CENSOR.....	56
2.11.4.	Analisi delle sequenze con il database CDD.....	56
2.11.5.	Analisi filogenetiche	56
	Capitolo terzo.....	59
	RISULTATI.....	59
3.1.	Valutazione generale delle sequenze ottenute	59
3.1.1.	Risultati della sequenza inserto del clone 1p2	61
3.1.2.	Risultati della sequenza inserto del clone 7p1	73
3.1.3.	Risultati della sequenza inserto del clone 59p1	83
	Capitolo quarto	90
	DISCUSSIONE	90
4.1.	Sequenza Ty3/Gypsy	91
4.2.	Sequenza DIRS.....	94
4.3.	Sequenza SAT2	95
	Bibliografia	98

Capitolo Primo

INTRODUZIONE

1.1. Scopo della tesi

Gli organismi eucariotici variano ampiamente per dimensione genomica e molte di queste variazioni possono essere spiegate sulla base della differenza dell'abbondanza del DNA ripetitivo (Charlesworth et al., 1994). La componente ripetitiva è costituita principalmente da DNA ripetuto in tandem, costituito da monomeri di lunghezza variabile a seconda che si tratti di minisatellite (tra 10 e 100 bp), microsatellite (tra 2 e 6 bp) o DNA satellite (in media 200 bp) (Biscotti et al., 2015) e da elementi trasponibili (TE), costituiti da sequenze con la capacità di spostarsi all'interno del genoma (Carducci et al., 2019; Makałowski et al., 2019). Inizialmente questa componente era stata considerata come “*junk DNA*”, ovvero “DNA spazzatura”, in quanto si considerava una porzione non-codificante e che quindi non avesse una funzione (Gregory, 2001). In seguito, diversi ruoli sono stati attribuiti a questo tipo di DNA tale da giustificarne la sua conservazione nel genoma degli organismi (Paço et al., 2019). Uno di questi è sicuramente il ruolo importante nella diversificazione degli organismi nel corso dell'evoluzione (Canapa et al., 2016; Gentles et al., 2005).

L'analisi del DNA ripetitivo è stata valutata in *Cynops orientalis*. Si tratta di una salamandra orientale, appartenente agli urodeli, che assieme ai gimnofioni e agli anuri appartiene al gruppo monofiletico dei Lissamphibia. Risulta importante studiare organismi come *C. orientalis* e altri anfibi perché possono essere utilizzati come animali modello per la loro fisiologia, diversità, filogenesi, per il loro sviluppo e per le spettacolari capacità di rigenerazione dei tessuti (Biscotti et al., 2020). Inoltre tra i vertebrati la *C. orientalis* possiede, come i dipnoi, un genoma di grandi dimensioni il cui sequenziamento non è ancora disponibile. Si tratta, infatti, di una taglia circa tredici volte più grande di quello umano. Per cui risulta molto interessante comprenderne la composizione. In particolare alcuni studi (Sun et al., 2012) hanno messo in evidenza l'alta percentuale di porzione ripetitiva, connessa ad un basso tasso di perdita del genoma nelle salamandre rispetto a uccelli e rettili (Gardner et al., 2020; J. Hanken & Wake, 1993).

Alla luce di ciò, è stata costruita una libreria genomica che permettesse di identificare per la prima volta elementi della porzione ripetitiva nel genoma di *C. orientalis*. In particolare sono state identificate tre sequenze per le quali sono state investigate l'organizzazione e la presenza in termini quantitativi nel genoma di *Cynops*. Attraverso analisi bioinformatiche, è stata verificata la diffusione di tali sequenze nel genoma di altri organismi ottenendo

informazioni sulla loro storia evolutiva. Inoltre, l'analisi sul trascrittoma ha permesso di verificare l'attività trascrizionale degli elementi isolati.

1.2. *Caudata*

I Caudata, anche detti urodeli, sono un gruppo di anfibi che include le salamandre e i tritoni. Fanno parte, insieme agli anuri e ai gimnofioni, del gruppo monofiletico dei Lissamphibia (FIGURA 1). I loro antenati comuni risalgono al Permiano, 300-250 milioni di anni fa (Sun et al., 2012). Gli urodeli si distinguono dagli altri due taxa, anuri e gimnofioni, per la morfologia: infatti sono caratterizzati da un corpo lacertiforme, dalla dimensione della testa ben proporzionata, dagli arti rugosi e con lunghezza simile tra loro e dalla coda permanente e lunga tanto quanto la loro testa e il corpo (Wake & Koo, 2018).

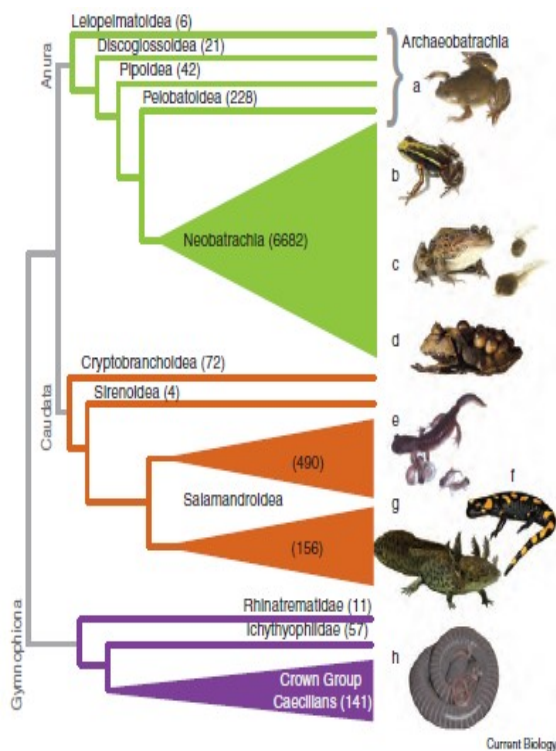


FIGURA 1 Relazione evolutiva tra i tre ordini degli anfibi (Wake & Koo, 2018).

La filogenesi dei Caudata è tuttora controversa. La classificazione ha fatto affidamento sull'indagine molecolare, affiancata a valutazioni sulla morfologia e al meccanismo di fecondazione (Faivovich et al., 2006). È evidente la distinzione dei seguenti cladi: Cryptobranchoidea, Sirenoidea e Salamandroidea.

In generale, questi organismi sono distribuiti geograficamente nelle regioni a Nord dell'Equatore. Per quanto riguarda la riproduzione sono ovipari semi-acquatici a fecondazione interna: i maschi producono degli spermatozoi, contenuti in strutture gelatinose chiamate spermatofore; durante il corteggiamento e l'accoppiamento i maschi lasciano una spermatofora sul terreno e le femmine la raccolgono attraverso le labbra cloacali, accumulando gli spermatozoi in un'area specifica della cloaca che si chiama spermateca (Buckley, 2012). Fanno eccezione i Cryptobranchoidea che effettuano fecondazione esterna.

A seguito della fecondazione si hanno diversi stadi di sviluppo: uno stadio larvale, in cui sono dotati di branchie e pinne, ed uno successivo di pedomorfosi o di neotenia, che prevedono il mantenimento di un fenotipo larvale anche se l'organismo ha raggiunto uno sviluppo sessuale (Denoël et al., 2005).

Una capacità molto affascinante e molto studiata degli urodela è la rigenerazione degli arti. Si tratta di un processo, presente già tra alcuni degli organismi primitivi, come idrozoi (Bode, 2003) e planarie (Rink, 2013), che consiste nel rigenerare parti del corpo, come ad esempio arti, coda, midollo spinale, mascella, retina, neuroni e cuore (Tang et al., 2017). In questo processo si distinguono tre stadi principali: 1) copertura della ferita: le cellule epidermiche circondanti la zona lesa, entro un giorno dall'amputazione, migrano verso la zona amputata e costituiscono un nuovo strato epidermico detto *wound epidermidis* (WE); 2) dedifferenziazione delle cellule circostanti: le cellule proliferano e formano uno strato mesenchimale, noto come blastema. Mentre WE si ispessisce fino a raggiungere la porzione apicale per formare l'*apical epidermal cap* (AEC); 3) differenziazione del blastema: favorisce la specializzazione dei tessuti e la formazione dell'arto lesa (Tang et al., 2017).

La taglia genomica dei Caudata è compresa tra i 10 ed i 120 Gb (Gregory, 2020; Wu, 2020). Si tratta di valori molto alti rispetto ai tetrapodi, ma una motivazione a questa condizione è data dall'abbondanza di sequenze non codificanti che vengono mantenute nel tempo e che vanno incontro a riarrangiamenti che ne favoriscono l'aumento di numero e la dispersione nel genoma (Sun et al., 2012). I riarrangiamenti, a cui si fa riferimento, sono caratterizzati da eventi di ricombinazioni intra- e inter-cromosomiche e da

meccanismi di inserzione e delezione (INDELs) che prevedono aggiunta e rimozione di sequenze. I loro effetti possono portare a diverse conseguenze tra cui: corretto impacchettamento del DNA (Biscotti et al., 2015), variazioni nella composizione genica (con possibilità di generare instabilità) (Hedges & Deininger, 2007) e nell'espressione genica, e meccanismi epigenetici (Paço et al., 2019). A seguito di ciò si ha un aumento della diversità intra-specifica della composizione genica dove le variazioni nel corso della selezione naturale possono essere perse o fissate (Lynch, 2007).

C'è, inoltre, una correlazione tra quantitativo di DNA, dimensione cellulare e dimensione corporea. Le cellule con genomi più grandi, infatti, richiedono tempi lunghi per la divisione cellulare, lento metabolismo basale e l'organismo è caratterizzato da uno sviluppo graduale e grandi dimensioni corporee. Questa condizione caratterizza alcune specie di salamandre che hanno perduto la capacità di compiere la metamorfosi e che rimangono nella loro condizione giovanile pur avendo conseguito la maturità sessuale (Gregory, 2001). Tuttavia molti studi non trovano riscontro con questa ipotesi, in quanto l'evoluzione ha permesso degli adattamenti che non consentono di spiegare linearmente la correlazione tra taglia genomica, tasso metabolico e dimensione corporea (Gardner et al., 2020). Un esempio è la salamandra del genere *Thorius* che ha un genoma abbondante, cellule di grandi dimensioni, ma è di taglia ridotta

(Hanken, 1983; Sessions & Larson, 1987). Questa condizione ha comportato un adattamento delle cellule nelle fasi di sviluppo e quindi della morfogenesi, della crescita e del raggiungimento della morfologia dell'adulto (Levy & Heald, 2016). Infatti, sempre nel *Thorius* si è verificato un adattamento morfologico che ha comportato una semplificazione del cervello e degli organi di senso, causata dall' adattamento alla riduzione delle dimensioni corporee (Hanken, 1983).

1.2.1. Cynops orientalis

L'organismo studiato in questa tesi è *Cynops orientalis*. Fa parte della famiglia Salamandridae e raggruppa insieme ai tritoni asiatici moderni (come per esempio *Laotriton*, *Pachytriton*, *Paramesotriton*), tanto che è normalmente distribuito lungo la riva Yangtze e nell'area adiacente, includente Henan, Hubei, Anhui, Jiangsu, Zhejiang, Jiangxi e le province di Fujian. È un organismo terrestre, ma alcune popolazioni riescono a stare sott'acqua (Wu, 2020).

Classificazione scientifica:

- Regno: Animalia
- Phylum: Chordata
- Subphylum: Vertebrata
- Classe: Sarcopterygii

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

Si tratta quindi di salamandre (*FIGURA 2*) la cui taglia corporea si aggira intorno agli 8 cm per le femmine ed intorno ai 6-6,5 cm per i maschi. Sono di colore nero, ma sono cosparse di macchioline rosse o arancione sulla zona ventrale. A questa colorazione è correlata una funzione di difesa nei confronti dei predatori.



FIGURA 2 Cynops orientalis (Wu, 2020).

La stagione riproduttiva inizia a marzo e termina a luglio, con un picco tra il tardo aprile e la metà di maggio. La temperatura adeguata dell'acqua è tra 15-

23°C. Dopo che il maschio approccia la femmina, la fiuta accuratamente secondo una metodica di corteggiamento che si conclude con la deposizione della spermatofora che la femmina raccoglie.

1.2.2. Taglia e composizione genomica dei Caudata

Gli anfibi urodeli sono di particolare interesse per il genoma di grandi dimensioni. Quello di *C. orientalis* è di 44,27 Gb ovvero circa 13 volte più grande di quello umano.

Il gigantismo genomico, dovuto quindi all'abbondanza di DNA, è una caratteristica che gli urodeli condividono con un altro gruppo dei vertebrati, ovvero i dipnoi. Questo forte aumento della taglia genomica è stato collegato in particolare alla porzione non codificante che nel corso dell'evoluzione viene mantenuta (Vignali & Nardi, 1996). È ormai noto che questa porzione è caratterizzata da DNA ripetitivo costituito da sequenze presenti in più copie nel genoma. Sulla base della loro distribuzione distinguiamo elementi interspersi, che comprendono DNA trasposoni e retrotrasposoni, ed elementi in tandem, che comprendono satelliti, minisatelliti e microsatelliti distinguibili sulla base della lunghezza e numero delle unità ripetute (Garrido-Ramos, 2017). È difficile definire le funzioni specifiche di questa componente del DNA, ma diversi studi hanno evidenziato che ha ruoli diversi come ad esempio nella compattazione della cromatina per la corretta formazione dei cromosomi, nella

regolazione genica, nella determinazione del corretto volume nucleare ma anche cellulare. Queste osservazioni giustificerebbero il suo mantenimento nel genoma e la sua propagazione nel corso dell'evoluzione (Vignali & Nardi, 1996).

L'abbondanza del DNA genomico delle salamandre le rende un esempio eclatante di paradosso del valore-C. Per valore-C si intende la quantità di DNA (espressa in picogrammi) contenuto in una cellula riferito al suo assetto aploide. Questo valore, inizialmente, era una chiave di lettura per spiegare la correlazione tra quantità di DNA e complessità dell'organismo. Infatti essendo i geni costituiti da DNA, maggiore era la complessità dell'organismo, maggiore era la quantità di geni che ci si aspettava e quindi maggiore doveva essere la quantità di DNA. Ma questa ipotesi non è sussistita per molto, perché vari studi hanno mostrato che non c'è correlazione tra le due componenti ed ecco perché si parla di paradosso del valore-C (Gregory, 2001). Una prima spiegazione venne data ammettendo che non tutto il DNA è fatto di geni ma anzi molto di questo è costituito da DNA non codificante ed in particolare da DNA ripetitivo. Tuttavia non fu facile capire le funzioni di questo DNA per molto tempo tanto che venne definito "*junk DNA*" e "*selfish DNA*" (Chalopin et al., 2015). Queste definizioni rendevano ancora più difficile capire perché questo DNA "inutile" potesse resistere ai vari passaggi evolutivi (Gregory, 2001). Oggi un numero

sempre maggiori di lavori evidenzia che questo DNA svolge diversi ruoli ma soprattutto ha un impatto sulla plasticità genomica, rendendo il DNA non più un'entità statica che viene trasmessa di generazione in generazione con alta fedeltà e riarrangiata solo da eventi di crossing-over durante la meiosi (Rubin et al., 2015). Quindi se così tanto DNA nei genomi degli urodela è stato preservato nel corso dell'evoluzione è perché la selezione naturale ne ha favorito il mantenimento, senza andare incontro a meccanismi di miniaturizzazione del genoma (Hanken & Wake, 1993), come nel caso degli uccelli o degli amnioti (Canapa et al., 2016).

1.2.2.1. DNA ripetitivo

Il DNA genomico presenta delle sequenze uniche, ma anche delle sequenze ripetute che sono presenti molte volte nel genoma e comprendono: sequenze ripetute in tandem e sequenze intersperse (*FIGURA 3*).

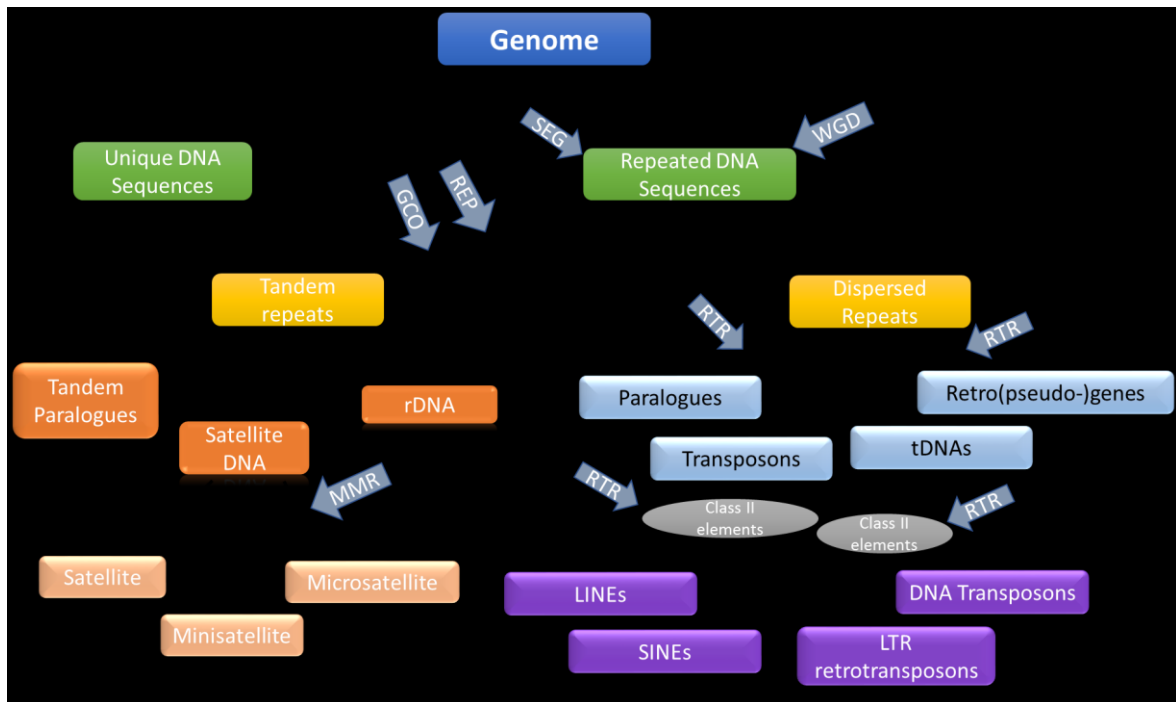


FIGURA 3 Tipologie di DNA ripetitivo nei genomi eucariotici e meccanismi di evoluzione. Due principali categorie di elementi ripetuti (in tandem e dispersi) vengono mostrate, a seguire le sottocategorie. Le frecce blu indicano i meccanismi molecolari coinvolti nella propagazione e l'evoluzione delle sequenze ripetute. REP, replication slippage; GCO, conversione genica; WGD, duplicazione dell'intero genoma; SEG, duplicazione segmentale; RTR, trascrizione inversa; TRA, trasposizione.

1.2.2.2. Sequenze ripetute in tandem

Le famiglie di DNA satellite (satDNA) sono sequenze organizzate in tandem altamente ripetute e non codificanti per proteine (Garrido-Ramos, 2017). Questi domini nella maggior parte dei casi sono presenti nell'eterocromatina e sono riconoscibili perché costituiti da centinaia a migliaia di copie della stessa sequenza motivo.

Si distinguono in satelliti, minisatelliti e microsatelliti e sono stati identificati nei principali gruppi tassonomici sia di specie vegetali che animali (Charlesworth et al., 1994).

I satelliti sono sequenze di DNA altamente ripetute, i cui monomeri misurano in media 200 bp. La lunghezza totale raggiunge diversi milioni di nucleotidi e le unità ripetute mostrano una grande variazione nel numero, creando array più o meno lunghi (Richard et al., 2008).

I minisatelliti sono sequenze di DNA composte da unità moderatamente ripetute la cui dimensione si aggira tra 10 bp e 100 bp.

I microsatelliti sono sequenze di DNA moderatamente ripetute composte da elementi di 2 - 6 bp (sempre minori di 9 bp).

Spesso la lunghezza della sequenza del satDNA è funzionale al ripiegamento del DNA intorno al nucleosoma, quindi da 150 bp a 180 bp (Biscotti et al., 2015). Si trovano frequentemente nella zona centromerica per l'evoluzione adattativa di istoni centromerici specifici che presentano un'alta affinità per le sequenze ripetute in tandem (Cooper & Henikoff, 2004; Hall, 2003; Widlund et al., 1999). In più alcune sequenze vengono trascritte generando *small interfering* RNA (siRNA) coinvolti in processi epigenetici di rimodulazione della cromatina e formazione di eterocromatina (Ugarkovic, 2005). Possono essere trascritti temporaneamente in particolari fasi di sviluppo o espressi differenzialmente nelle cellule, tessuti e organi differenti, come nel caso del tritone *Notophthalmus viridescens* (Epstein et al., 1986).

Molti studi hanno collegato l'origine e la formazione di satDNA a partire da sequenze di elementi trasponibili, dato che in alcuni casi si ha similarità tra le due tipologie di sequenze (Paço et al., 2019), basti pensare alle sequenze Alu o LINE1 che nel genoma dei mammiferi favoriscono inserzioni di satDNA, caratterizzate da lunghe sequenze di poli-A, provocando un'alterazione dell'architettura genetica ed epigenetica del genoma (Grandi et al., 2013). A questo consegue che il satDNA va incontro ad una riorganizzazione della sua composizione nucleotidica che porta anche ad una variazione del numero di ripetizioni che può provocare un'instabilità del genoma. Come nel caso dell'ataxia di Friedreich (Montermini et al., 1997) dove l'aumento delle ripetizioni dei microsatelliti, localizzati all'interno delle sequenze Alu nel primo introne del gene della frataxina, provoca una mutazione che porta ad una malattia autosomica recessiva neurodegenerativa.

Oltre alla classificazione per la lunghezza è possibile fare ulteriori distinzioni sulla base della dimensione della sequenza motivo, della lunghezza totale dell'array e della presenza nel genoma (*FIGURA 4*). È possibile notare come nel genoma siano ampiamente diffusi, anche se non facilmente riconoscibili, i microsatelliti e che i satelliti sono meno presenti in quanto generano instabilità.

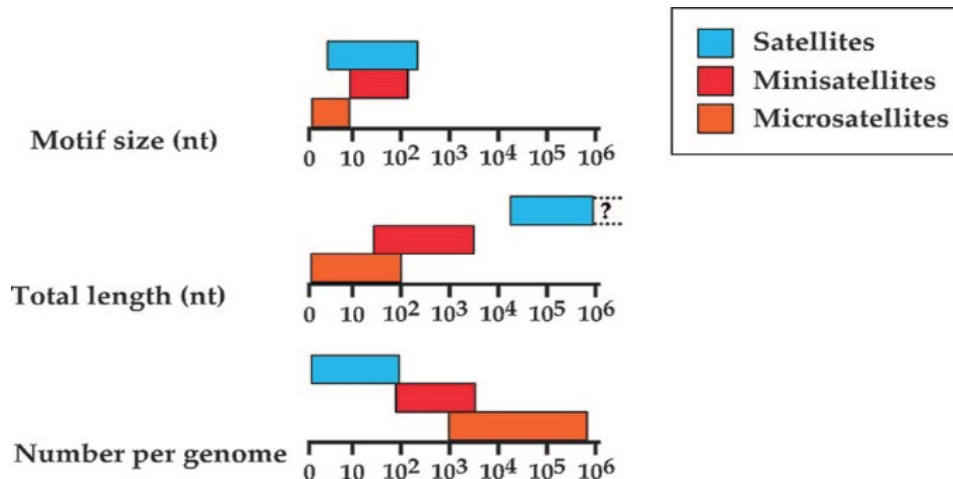


FIGURA 4 Dimensioni del motivo, lunghezza e abbondanza di sequenze ripetute in tandem (Richard et al., 2008).

1.2.2.3. Sequenze ripetute intersperse

Le sequenze ripetute intersperse sono sequenze presenti in più copie e distribuite casualmente nel genoma, infatti non sono raggruppate o organizzate in piccoli gruppi come le sequenze ripetute in tandem (Paço et al., 2019). Insieme a queste ultime hanno un importante impatto nel mantenimento della stabilità del genoma durante i processi di mitosi e meiosi, oltre che durante i meccanismi evolutivi di duplicazione segmentale o dell'intero genoma (Charlesworth et al., 1994; Richard et al., 2008). Sono presenti a livello dell'eterocromatina centromerica e telomerica nel caso in cui sono associati ai satDNA, ma anche all'interno di sequenze introniche e nelle regioni di regolazione della trascrizione (Biscotti et al., 2015).

Le sequenze ripetute intersperse includono i geni paraloghi e i retropseudogeni, ma la componente più abbondante, ubiquitaria e diversificata sia nelle piante che negli animali è quella degli elementi trasponibili (TE).

Gli elementi trasponibili sono frammenti di DNA in grado di compiere trasposizione, un meccanismo che consta nell'abilità di spostarsi all'interno di un genoma, in alcuni casi creando nuove copie di loro stessi durante lo spostamento (Sotero-Caio et al., 2017).

La classificazione dei TE si basa sul diverso meccanismo di trasposizione (Finnegan, 1989) e ci sono due classi: retrotrasposoni (classe I) e trasposoni a DNA (classe II) (TABELLA 1).

TABELLA 1. Classificazione degli elementi trasponibili eucariotici (Wicker et al., 2008).

Class	Order	Superfamily	Phylogenetic distribution	
Class I (retrotransposons)	LTR	Copia	Plants, metazoans, fungi	
		Gypsy	Plants, metazoans, fungi	
		Bel-Pao	Metazoans	
		Retrovirus	Metazoans	
		ERV	Metazoans	
	DIRS	DIRS	Plants, metazoans, fungi	
		Ngaro	Metazoans, fungi	
		VIPER	Trypanosomes	
	PLE	Penelope	Plants, metazoans, fungi	
	LINE	R2	Metazoans	
		RTE	Metazoans	
	SINE	Jockey	Metazoans	
		L1	Plants, metazoans, fungi	
		tRNA	Plants, metazoans, fungi	
		7SL	Plants, metazoans, fungi	
5S		Metazoans		
Class II (DNA transposons) Subclass 1	TIR	Tc1-Mariner	Plants, metazoans, fungi	
		hAT	Plants, metazoans, fungi	
		Mutator	Plants, metazoans, fungi	
		Merlin	Metazoans	
		Transib	Metazoans, fungi	
		P	Plants, metazoans	
		PiggyBac	Metazoans	
		PIF-harbinger	Plants, metazoans, fungi	
		CACTA	Plants, metazoans, fungi	
		Crypton	Crypton	Fungi
	Class II (DNA transposons) Subclass 2	Helitron	Helitron	Plants, metazoans, fungi
		Maverick	Maverick	Metazoans, fungi

Il meccanismo utilizzato dagli elementi della classe I è simile ad un “copia ed incolla”, in cui a partire dal DNA genomico si ottiene un RNA intermedio che poi viene retrotrascritto in una nuova copia di DNA da parte di una

retrotrascrittasi (RT), codificata dal TE. Questa copia viene trasferita ed integrata nel sito donatore per mezzo di una integrasi (INT) (Wicker et al., 2009). Questa classe include cinque ordini, che possono essere ulteriormente classificati in sequenze *Long Terminal Repeats* (LTR) e sequenze *non-Long Terminal Repeats* (non-LTR). La differenza tra i due sott'ordini è data dalla presenza di estremità terminali ripetute con una lunghezza variabile tra 250-600 bp. Le sequenze LTR si suddividono in tre superfamiglie presenti nei vertebrati: *Ty1/Copia*, *Ty3/gypsy-like* e *BEL/Pao*. Le sequenze non-LTR non presentano estremità terminali ripetute ed includono principalmente le sequenze *Long INterspersed Nuclear Elements* (LINE) e *Short INterspersed Nuclear Elements* (SINE) (Carducci et al., 2019) (FIGURA 5).

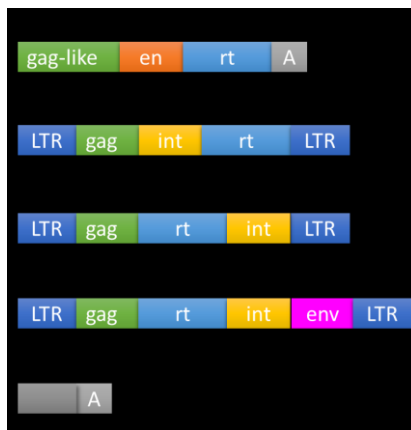


FIGURA 5 Principali differenze strutturali tra TE.

I retrotrasposoni LTR sono meno abbondanti negli animali e presentano origine retrovirale. La loro lunghezza va da poche centinaia di basi ad eccezionalmente 25 kb. Per l'integrazione producono una *Target Sequence Duplication* (TSD)

lunga 2-6 bp. I retrotrasposoni LTR presentano due regioni, tipiche dei retrovirus: *gag* per la produzione di strutture proteiche per particelle virali e *pol* per le proteine di replicazione. A differenza dei retrovirus non hanno *env* che codifica per la produzione di proteine di membrana e quindi non hanno capacità di passare da una cellula all'altra e non riescono ad infettarle (Benachenhou et al., 2013; Dodonova et al., 2019; Malik et al., 2000; Ribet et al., 2008). Le due superfamiglie maggiormente presenti sono *Gypsy* e *Copia* e differiscono tra di loro per l'ordine della RT e INT nella regione *pol* (Wicker et al., 2009).

Oltre ai TE di classe I, sono stati caratterizzati due altri gruppi: *Dictyostelium intermediate repeat sequence* (DIRS) e *Penelope-like*. Gli elementi DIRS sono lunghi circa 5 kb (Cappello et al., 1985). La sequenza è caratterizzata da estremità ripetute invertite (TIR) non identiche, come nel caso del *Dictyostelium*, o anche da estremità *split direct repeats* (SDRs), in cui le sequenze ripetute terminali sono adiacenti l'una all'altra nella regione interna complementare (ICR), come nel caso di *Panagrellus redivivus*. Mantengono la normale composizione di un elemento LTR ma al posto del gene per l'integrasi hanno un gene per la tirosina-ricombinasi. Quest'ultima promuove meccanismi di ricombinazione sito-specifica tra sequenze di DNA simili o identiche (Cappello et al., 1985; Muszewska et al., 2013). Le sequenze *Penelope-like* sono presenti tra amebe e funghi ma anche in vertebrati che non siano

mammiferi. Tra gli invertebrati sono stati studiati in *Drosophila virilis*, la cui presenza provoca disgenesi, caratterizzata da trasposizione simultanea di TE, appartenenti a diverse famiglie, nella prole con incroci disgenici. La sequenza *Penelope-like* codifica per dominio RT, simile alla sequenza della telomerasi, ed endonucleasi (EN). Le estremità sono *LTR-like*, con orientamento diretto o inverso e con funzione intronica (Metcalf & Casane, 2013).

I retroelementi non-LTR, presenti prevalentemente negli eucarioti, sono gli elementi LINE; si tratta di sequenze lunghe 8-9 kb, possiedono un dominio legante RNA e due *Open Reading Frame* (ORF) che trascrivono per retrotrascrittasi ed endonucleasi, rimanendo quindi autonomi nella replicazione (Gentles et al., 2005). Mentre gli elementi SINE sono meno presenti e incapaci di replicare autonomamente perché derivano dall'evoluzione di geni 7SL e tRNA, e quindi sono il frutto di retrotrasposizione accidentale della Polimerasi III (Pol III) di sequenze lunghe al massimo 1000 bp (Makałowski et al., 2019). Questi TE sono organizzati e distribuiti nel genoma, anche se poco rappresentati, considerando i risultati degli assemblaggi genomici di alta qualità. Molte di queste sequenze non sono presenti per la perdita di DNA per ricombinazione ectopica o per l'alto tasso di sostituzione, come in *Anolis carolinensis* (Tollis & Boissinot, 2013).

Gli elementi TE della classe II sono abbastanza distribuiti in tutti gli eucarioti. Si tratta di elementi, riscontrati già nei procarioti, per la presenza di strutture semplici come le sequenze di inserzione (IS) o di strutture più complesse. Vengono ulteriormente distinti in due sottoclassi sulla base del numero di filamenti di DNA che vengono tagliati durante la trasposizione, ma non per generare l'RNA intermedio come nelle sequenze LTR.

La sottoclasse I prevede il meccanismo “taglia ed incolla”, favorito dalla presenza da sequenze terminali ripetute invertite (TIR) con lunghezza variabile. Questi TE possono aumentare il loro numero tramite trasposizione durante la replicazione del cromosoma da una posizione ad un'altra. La trasposasi riconosce le sequenze TIR e taglia entrambi i filamenti in ogni punto terminale.

La sottoclasse II comprende TE a DNA che seguono un meccanismo di trasposizione diverso che non richiede il taglio del doppio filamento, ma anzi seguono dei meccanismi di replicazione a *rolling circle* come nel caso degli *Helitron* (Grau et al., 2014).

La considerazione degli elementi trasponibili è cambiata nel tempo: inizialmente erano ritenuti una porzione genomica priva di funzione, essendo considerati parte del *selfish* o *junk* DNA (Chalopin et al., 2015). Successivamente si è compreso la loro importanza nei processi evolutivi ed in

particolare nella “*molecular domestication*”, ovvero nell’essere cooptati per creare nuovi geni codificanti proteine durante l’evoluzione della maggior parte degli eucarioti. Quindi hanno un ruolo chiave nell’evoluzione e nella diversificazione dei vertebrati. Basti considerare geni derivati da retroelementi LTR, coinvolti nella formazione della placenta, funzioni cognitive nel cervello e immunità o anche nella proliferazione cellulare, apoptosi e nel cancro (Naville et al., 2016).

L’ubiquità di TE pone innumerevoli sfide per il mantenimento dell’integrità genomica, dato che porta a riarrangiamenti cromosomici durante ricombinazione ectopica TE-TE ma anche durante il processo di trasposizione stesso (Hedges & Deininger, 2007). In altri casi altera la regolazione dell’espressione genica, provocando diversificazione intra- e inter-specie attraverso interruzione degli esoni, con induzione di mis-splicing, prematura poliadenilazione dei trascritti e alterazione trascrizionale dei geni (Kordis, 2010). Quindi per l’elevato numero di elementi mobili ed i loro effetti si conferisce plasticità al genoma (Bonett et al., 2009).

Capitolo secondo

MATERIALI E METODI

2.1. Campionamento

In questo studio è stato utilizzato l'esemplare *C. orientalis* acquistato dal personale del laboratorio di Filogenesi molecolare della Facoltà di Scienze, Università Politecnica delle Marche, Ancona (Italia). L'esemplare è stato sacrificato e i suoi tessuti sono stati conservati in *RNAlater® Tissue Collection* (Applied Biosystems). L'*RNAlater* è un reagente acquoso e non tossico, utilizzato per la conservazione dei tessuti: riesce a penetrare rapidamente nel tessuto andando a stabilizzare e a proteggere gli acidi nucleici presenti nei campioni freschi, ed in particolare quelli ribonucleici che per la loro struttura a singolo filamento sono più facilmente soggetti a degradazione. L'*RNAlater* elimina la necessità di processare immediatamente i campioni di tessuto o di congelarli in azoto liquido per l'elaborazione successiva.

2.2. Estrazione del DNA totale mediante alcool isoamilico e cloroformio (CIA)

Il DNA di *C. orientalis* è stato estratto prelevando circa 1 grammo di tessuto muscolare. Il tessuto è stato reso in pezzetti più piccoli mediante l'uso di forbicine, poi si è proceduto con il seguente protocollo:

1. omogeneizzare il materiale con potter a mano in 5 ml di tampone contenente: 0,1 M di EDTA a pH 8,0 e 0,15 M di NaCl;
2. trasferire il materiale in una nuova provetta;
3. aggiungere 1/10 del volume di SDS (Sodio Dodecil Solfato al 10%) e 0,1 mg/ml di proteinasi K. In questa fase si verificherà la lisi delle cellule attraverso la solubilizzazione delle membrane e la degradazione delle proteine permettendo quindi l'estrazione degli acidi nucleici;
4. incubare a 37°C in bagnetto per tutta la notte;
5. aggiungere 1/10 del volume di perclorato di sodio 5 M (o 1/10 del volume di sodio acetato 3 M) ed un volume di cloroformio ed alcool isoamilico in rapporto 24:1;
6. estrarre agitando delicatamente per 10 minuti;
7. centrifugare a 7.000 x g per 15 minuti;
8. recuperare la fase superiore acquosa aspirando delicatamente con una pipetta pasteur, evitando di aspirare le proteine che sono precipitate a livello dell'interfaccia fra le due fasi (*FIGURA*);

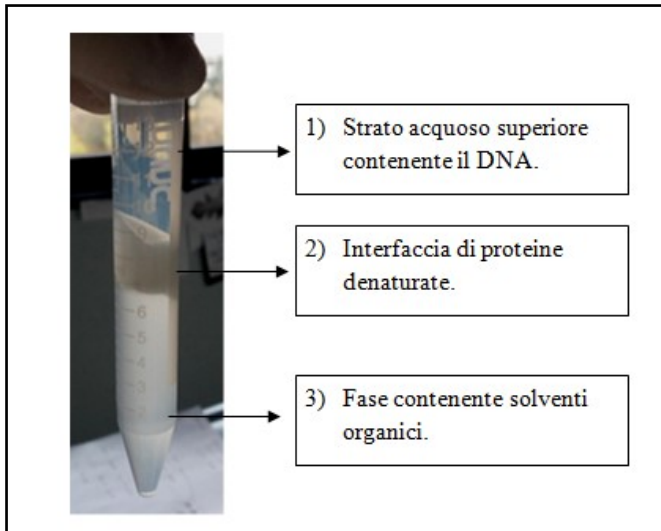


FIGURA 6 Fasi osservabili durante l'estrazione del DNA.

9. precipitare il DNA aggiungendo 2 volumi di etanolo assoluto (EtOH) freddo mantenuto a -20°C ;
10. invertire 2 - 3 volte la provetta: formazione di una nuvola di DNA che viene recuperata e lavata con etanolo freddo 70% (*FIGURA 77*);



FIGURA 7 Nuvole di DNA che si formano in seguito all'aggiunta di etanolo assoluto.

11. asciugare il DNA estratto sotto cappa o portarlo a secco con una speed-vac e poi risospendere in TE (10 mM Tris HCl, 1 mM EDTA pH 8,0);
12. effettuare una lettura al Qubit™, per valutare la concentrazione e la qualità del DNA estratto.

2.3. Quantizzazione del DNA estratto con Qubit™

Il fluorimetro Qubit™ è un dispositivo di nuova generazione per misurare accuratamente quantità di DNA, RNA e proteine. È in grado di rilevare coloranti fluorescenti che si legano specificatamente a target d'interesse, anche in campioni che richiedono alta sensibilità.

Con il kit *Invitrogen double strand DNA high sensitivity* è stato possibile quantizzare le concentrazioni di DNA dei campioni, secondo il seguente protocollo:

1. preparare due eppendorf per gli standard ed una per il campione;
2. preparare la *working Solution Qubit* seguendo la diluizione 1:200 del *Qubit® dsDNA HS Reagent* nel *Qubit® dsDNA HS Buffer*. Sono necessari 200 µl di *Working Solution* per ogni standard e per ogni campione;
3. preparare le provette come di seguito:

	Provetta per lo standard	Provetta per il campione
<i>Volume di working Solution</i>	190 μ l	198 μ l
<i>Volume di standard</i>	10 μ l	-
<i>Volume di campione</i>	-	2 μ l
<i>Volume totale</i>	200 μ l	200 μ l

4. vortexare le provette per 2-3 secondi;
5. incubarle per 2 minuti a temperatura ambiente;
6. inserire le provette nel fluorimetro e annotare il risultato della lettura.

2.4. Costruzione della libreria genomica

Una libreria genomica rappresenta una collezione di frammenti di DNA rappresentativi del genoma di un organismo.

Per la costruzione di una libreria genomica è necessario frammentare il DNA genomico, precedentemente estratto, tramite l'utilizzo di un enzima di restrizione che taglia frequentemente il genoma in esame, ma anche il sito di clonaggio multiplo di un vettore nel quale viene inserito il frammento. Per cui ogni clone conterrà un frammento di DNA e ogni colonia derivante da quel clone sarà portatrice di quel preciso frammento.

Affinché una libreria sia utile, è necessario che contenga almeno una copia di ogni frammento generato in modo da rappresentare tutto il genoma.

2.4.1. Trattamento con RNasi

Il DNA estratto è stato trattato con RNasi per eliminare l'eventuale presenza di RNA nel campione. A tale scopo è stata allestita la seguente reazione:

- 100 µl di DNA (10 ng/ml) di *C. orientalis*;
- 1x di Reaction Buffer;
- 1 µl di RNasi (4 mg/ml).

Tale reazione è stata incubata a 37°C per circa 2 ore.

2.4.2. Digestione del DNA mediante endonucleasi di restrizione

Il DNA del *C. orientalis* è stato frammentato utilizzando l'endonucleasi di restrizione *Pst*I. Tale enzima produce estremità protrudenti che favoriscono la ligazione dei frammenti ottenuti con il vettore plasmidico. La reazione di digestione è stata così preparata:

REAGENTI	QUANTITA'
DNA di <i>C. orientalis</i> (concentrazione madre =191,0 ng/ µl)	3 µl
<i>Pst</i>I (10 U/µl)	1 µl
Restriction Enzyme 10X Buffer	5 µl
Albumina sierica bovina (BSA) (10 µg/µl)	0,5 µl
Acqua deionizzata sterile	40,5 µl
Volume finale	50 µl

Tale reazione è stata incubata a 37°C per tutta la notte. La frammentazione del genoma è stata verificata con corsa elettroforetica su gel di agarosio 1,5%. Successivamente il materiale è stato precipitato con sodio acetato ed etanolo.

2.4.3. Selezione dei frammenti di interesse da clonare e purificazione degli acidi nucleici da gel di agarosio

Per selezionare i frammenti di interesse è stata effettuata una corsa elettroforetica del DNA di *C. orientalis*, digerito con *Pst*I, con il marcatore di peso molecolare AmpliSize 50 - 2.000 bp ladder (BIO-RAD) avente bande di lunghezza nota. L'elettroforesi è stata condotta su gel di agarosio al 2% in tampone TBE 1x (45 mM Tris borato; 1 mM EDTA pH 8,0) su un apparato elettroforetico orizzontale *Sub-Cell Systems* (BIO-RAD). I frammenti tra 100 e 1.500 bp sono stati isolati e purificati da gel utilizzando il kit *PureLink*TM (ThermoFisher) con il seguente protocollo:

1. tagliare la banda di interesse dal gel con un bisturi ed inserirla in una provetta vuota precedentemente pesata;
2. aggiungere una quantità di *Gel Solubilization Buffer* (L3) pari a 3 volte il peso in grammi della banda stessa;
3. incubare il campione a 50°C per 10 minuti ed invertire la provetta delicatamente ogni 3 minuti;

4. incubare la provetta ulteriori 5 minuti per favorire la completa dissoluzione del gel;
5. trasferire il gel dissolto nell'apposita colonna di purificazione costituita da una membrana di silice che consente di trattenere il DNA in presenza di elevate concentrazioni saline;
6. centrifugare per 1 minuto a 12.000 x g a temperatura ambiente, eliminando successivamente il liquido presente sul fondo del "wash tube";
7. aggiungere alla colonna di purificazione 500 µl di Wash Buffer (W1) che consente di rimuovere le rimanenti tracce di agarosio legate al campione;
8. centrifugare per 1 minuti a 12.000 x g;
9. rimuovere il liquido presente nel "wash tube" e centrifugare ulteriormente per 3 minuti alla massima velocità;
10. posizionare la colonnina cromatografica su una nuova provetta ed aggiungere 30 µl di Elution Buffer (E1);
11. incubare per 1 minuto a temperatura ambiente e infine centrifugare per 1 minuto a 12.000 x g.

2.4.4. Clonaggio

Per la creazione delle librerie di *C. orientalis* è stato utilizzato il vettore *pGEM®-T Easy* (FIGURA8) del kit *pGEM®-T and pGEM®-T Easy Vector*

Systems (PROMEGA). Il kit comprende due vettori pGEM®-T e pGEM®-T Easy. L'unica differenza tra i due vettori è la presenza in pGEM®-T Easy di ulteriori 15 bp nel sito di clonaggio multiplo che offrono una più ampia scelta di enzimi di restrizione da poter utilizzare rispetto a pGEM®-T. Il vettore include:

- gene codificante la resistenza all' ampicillina per la selezione in *E. coli*;
- origine di replicazione di f1;
- siti per i primer M13, utili in fase di sequenziamento;
- sito di clonaggio multiplo.

Inoltre, il vettore pGEM®-T Easy presenta i promotori dell'RNA polimerasi T7 e SP6 (per la trascrizione *in vitro*) che fiancheggiano il sito di clonaggio multiplo, il quale si trova all'interno della regione che codifica per l' α -peptide dell'enzima β -galattosidasi (gene *lacZ*) e che viene utilizzato per la selezione bianco-blu dei cloni ricombinanti.

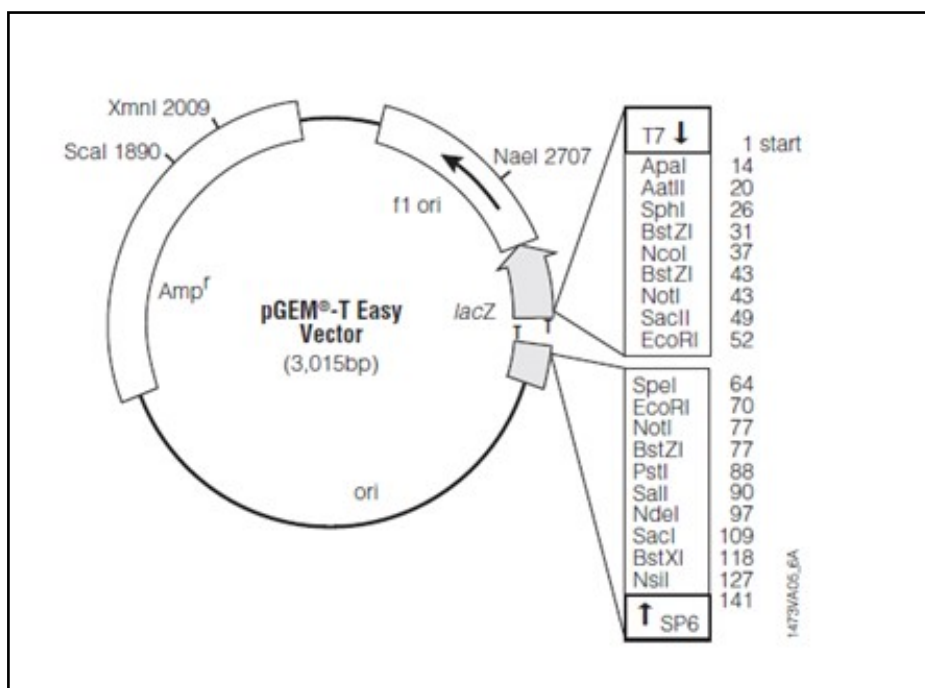


FIGURA 8 Mappa del plasmide pGEM®-T EASY.

Se la cellula batterica viene trasformata dal vettore non ricombinante (privo dell'inserito), in presenza di un induttore artificiale come l'IPTG (analogo del lattosio) e di X-Gal (substrato cromogeno della β -galattosidasi), il gene lacZ produce la β -galattosidasi che scinde l'X-Gal producendo colonie batteriche di colore blu. Se invece, nel gene lacZ del plasmide è stato inserito del DNA estraneo, quindi il vettore è diventato vettore ricombinante, le cellule batteriche trasformate non producono più la β -galattosidasi attiva (per interruzione del gene lacZ) e pertanto verranno prodotte colonie batteriche bianche (FIGURA).

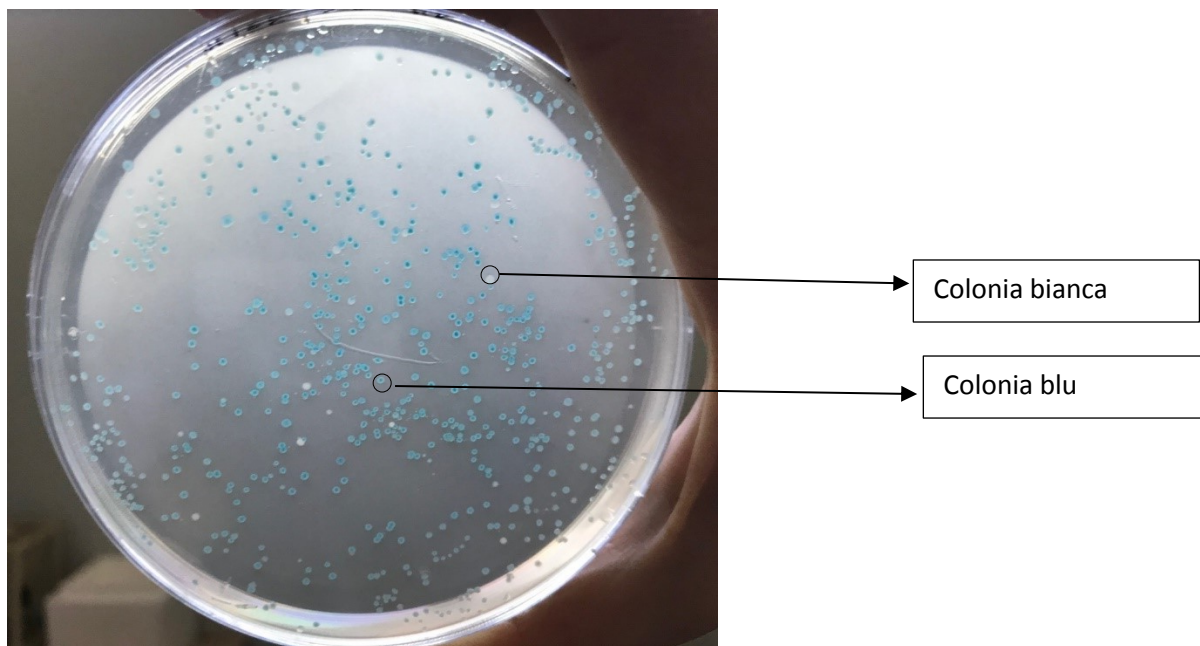


FIGURA 9 Valutazione delle piastre dopo la crescita dei cloni mediante selezione bianco/blu.

2.4.5. Ligazione

Il frammento di interesse viene inserito in un vettore plasmidico tagliato con lo stesso enzima di restrizione, mediante la seguente reazione di ligazione:

REAGENTI	QUANTITA' (μ l)
2X Rapid Ligation Buffer, T4 DNA Ligase	7,5 μ l
pGEM®-T Easy Vector (50 ng)	1,0 μ l
Inserito di DNA tagliato	5,0 μ l
T4 DNA Ligase (3 Weiss units/μl)	1,0 μ l
Acqua nuclease-free fino a un volume finale di	15 μl

Per ottenere il massimo numero di trasformanti, la reazione è stata incubata a 4°C per tutta la notte.

2.4.6. Trasformazione mediante shock termico

Al fine di inserire il prodotto della ligazione nelle cellule competenti *E. coli* JM109, è stato effettuato lo shock termico il quale genera la formazione di pori a livello della parete e della membrana cellulare che ne consentono il passaggio.

La procedura di trasformazione è stata svolta seguendo questo protocollo:

1. scongelare in ghiaccio le cellule *E. coli* JM109;
2. trasferire le cellule in una provetta di polipropilene sterile ed aggiungere 2 μ l della reazione di ligazione;
3. incubare la provetta in ghiaccio per 20 minuti;
4. successivamente provocare shock termico a 42°C per 45 - 50 secondi;
5. trasferire la provetta rapidamente in ghiaccio per 2 minuti;
6. aggiungere alla provetta contenente le cellule trasformate 950 μ l di S.O.C. Medium (2% tryptone, 0,5% estratto di lievito, NaCl 10 mM, KCl 2,5 mM, MgCl₂ 10 mM, MgSO₄ 10 mM, glucosio 20 mM);
7. incubare a 37°C in agitazione a 190 rpm per 1 ora e mezza;
8. centrifugare le cellule a 1.000 x g per 10 minuti, eliminando il sovrantante;
9. risospendere le cellule in 200 μ l di S.O.C. medium.

2.4.7. Semina

La semina delle cellule trasformate viene effettuata su piastre contenenti terreno LB medium/Ampicillina (agar 15 g/l, tryptone 10 g/l, estratto di lievito

5 g/l, NaCl 10 g/l, pH 7.0; 1 ml di Ampicillina per 500 ml di LB medium), procedendo nel seguente modo:

1. porre sulle piastre contenenti terreno LB medium/Ampicillina 100 µl di IPTG (0,1 M) (isopropil-β-D-1-tiogalattopiranoside) e 20 µl di X-Gal (50 mg/ml) (bromo-cloro-indolil-galattopiranoside), distribuendoli in maniera uniforme sulla superficie con una spatola;
2. incubare le piastre a 37°C per 30 minuti;
3. porre sulle piastre di semina 100 µl di cellule trasformate distribuendole uniformemente sulla superficie del terreno;
4. infine, incubare le piastre a 37°C per tutta la notte.

Lo stesso procedimento è stato effettuato per le diluizioni 1:50 e 1:100.

2.5. Screening della libreria

2.5.1. Marcatura del DNA genomico di *C. orientalis*

La libreria ottenuta è stata analizzata utilizzando come sonda il DNA genomico di *C. orientalis*. A tale scopo si è proceduto degradando 1 µg di DNA a 110°C per 4 minuti in autoclave. Per la marcatura del DNA genomico è stato utilizzato il kit *DIG High Prime DNA Labeling and Detection Starter Kit I (Roche)*. Il DNA è stato denaturato a 96°C per 10 minuti in termociclatore, velocemente raffreddato in ghiaccio e marcato aggiungendo 4 µl della miscela *DIG-High Prime (Roche)*. La reazione è stata incubata a 37°C per tutta la notte ed infine bloccata mediante un passaggio a 65°C per 10 minuti.

2.5.2. Quantificazione della sonda e determinazione dell'efficienza della reazione di marcatura

Per verificare l'efficienza della marcatura sono state effettuate una serie di diluzioni seriali di DNA come riportato nella tabella sottostante.

Provetta	DNA (μ l)	Dalla provetta	DNA dilution Buffer	Diluizione	Conc. finale
1	-	Diluzione originale			1 ng/ μ l
2	2	1	198 μ l	1:100	10 pg/ μ l
3	15	2	35 μ l	1:3,3	3 pg/ μ l
4	5	2	45 μ l	1:10	1 pg/ μ l
5	5	3	45 μ l	1:10	0,3 pg/ μ l
6	5	4	45 μ l	1:10	0,1 pg/ μ l
7	5	5	45 μ l	1:10	0,03 pg/ μ l
8	5	6	45 μ l	1:10	0,01 pg/ μ l
9	0	-	50 μ l	-	0

Una volta effettuate tutte le diluzioni, il protocollo prevede i seguenti passaggi:

1. spottare 1 μ l delle diluizioni dalla 2 alla 9 di sonda da controllare e di sonda di riferimento su una membrana di nylon carica positivamente;
2. fissare il DNA per 30 minuti a 120°C;
3. trasferire la membrana in 20 ml di acido maleico Buffer (0,1 M acido maleico, 0,15 M NaCl, pH 7,5);
4. agitare per 2 minuti a 15 - 20°C;
5. incubare per 30 minuti in 5 ml di *Blocking Solution*;
6. incubare per 30 minuti in 5 ml di *Antibody Solution*;
7. lavare in 10 ml di *Washing Buffer* (0,1 M acido maleico, 0,15 M NaCl pH 7,5; 0,3% Tween 20) per 15 minuti e ripetere questo passaggio;

8. incubare in 10 ml di *Detection Buffer* (0,1 M Tris HCl, 0,1 M NaCl; pH 9,5) per 5 minuti;
9. incubare al buio per 30 minuti in 2 ml di *Color Substrate Solution* (40 µl di NBT/BCIT stock Solution in 2 ml di *Detection Buffer*);
10. lavare la membrana per 5 minuti in 50 ml di acqua bidistillata.

2.5.3. Trasferimento delle colonie della libreria su membrana

Per poter individuare quali delle colone di *C. orientalis* contenessero all'interno del vettore *pGEM®-T Easy* frammenti di DNA ripetuto, si è proceduto con la crescita delle colonie in LB liquido e il trasferimento di 1 µl di ognuna su una membrana di nylon posizionata su una piastra Petri con LB solido e lasciata incubare tutta la notte in termostato a 37°C. Il DNA è stato denaturato e fissato attraverso i seguenti passaggi:

1. porre della carta assorbente sul fondo di 5 piastre Petri;
2. saturare ogni piastra rispettivamente con una soluzione di SDS 10%, una *Denaturing Solution* (1,5 M NaCl e 0,5 M NaOH), una *Neutralizing Solution* (1,5 M NaCl e 0,5 M Tris HCl pH 7,4) ed infine 20× SSC;
3. eliminare gli eccessi di liquido ed evitare la formazione di bolle;
4. incubare per 3 minuti la membrana su una piastra Petri contenente carta imbevuta di una soluzione di SDS 10%;

5. trasferire la membrana su una seconda piastra contenente la *Denaturation Solution* (1,5 M NaCl e 0,5 M NaOH) ed incubare per 5 minuti;
6. trasferire la membrana su una terza piastra contenente *Neutralizing solution* (1,5 M NaCl e 0,5 M Tris HCl pH 7,4) ed incubare per 5 minuti;
7. ripetere il passaggio precedente;
8. trasferire la membrana su una quarta piastra con soluzione di SSC 20× ed incubare per 5 minuti;
9. asciugare la membrana per 30 minuti a temperatura ambiente;
10. fissare il DNA per 30 minuti a 120°C in stufa.

2.5.4. Ibridazione del genoma marcato sulla membrana

La procedura di ibridazione del genoma marcato sulla membrana precedentemente ottenuta prevede:

1. pre-riscaldare un appropriato volume di *DIG Easy Hyb* (10 ml/100 cm² di membrana) alla temperatura di ibridazione (42°C);
2. pre-ibridizzare la membrana con 10 ml/100 cm² di membrana di *DIG Easy Hyb* per 30 minuti a 42°C agitando delicatamente;
3. denaturare la sonda marcata con digossigenina (circa 25 ng/ml di *DIG Easy Hyb*) ponendola a 96°C in termociclatore per 5 minuti e poi rapidamente in ghiaccio;

4. eliminare la soluzione di pre-ibridazione e aggiungerne altra fresca secondo le quantità 3,5 ml/100 cm² di membrana;
5. aggiungere la sonda denaturata;
6. incubare per tutta la notte agitando gentilmente a 42°C.

2.5.5. Lavaggi stringenti post-ibridazione

Dopo l'ibridazione vengono effettuati una serie di lavaggi sotto costante agitazione:

1. lavare 2 volte per 5 minuti in una soluzione di SSC 2× e SDS 0,1% a 15-25°C;
2. lavare 2 volte per 15 minuti in una soluzione di SSC 0,5× e 0,1% SDS a 65-68°C.

2.5.6. Determinazione immunologica

Per la determinazione immunologica si procede secondo i seguenti passaggi:

1. risciacquare la membrana per 1-5 minuti in *Washing Buffer* (0,1 M acido maleico, 0,15 M NaCl pH 7,5, 0,3% Tween 20);
2. incubare in agitazione per 30 minuti in 20 ml di *Blocking Solution* 1×;
3. incubare in agitazione per 30 minuti in 20 ml di *Antibody Solution*;
4. lavare 2 volte in 20 ml di *Washing Buffer* per 15 minuti in agitazione;
5. aggiungere 20 ml di *Detection Buffer* (0,1 M Tris HCl pH 9,5, 0,1 M NaCl) per 2-5 minuti in agitazione;

6. incubare la membrana in 10 ml di *Color Substrate Solution* (200 µl di *NBT/BCIP stock Solution* in 10 ml *Detection Buffer*) preparata fresca e conservare al buio per tutta la notte;
7. bloccare la reazione lavando la membrana per 5 minuti con 50 ml di acqua bidistillata sterile.

2.6. *Screening delle colonie di interesse mediante reazione a catena della polimerasi (PCR)*

Le colonie di interesse, ossia quelle che hanno mostrato un intenso segnale dopo la determinazione immunologica sulla membrana, sono state amplificate per mezzo della reazione a catena della polimerasi (PCR), al fine di verificare anche la corretta integrazione dell'inserito all'interno del vettore.

Intorno alla metà degli anni '80, il biochimico Kary Mullis (premio Nobel per la chimica nel 1993) mise a punto un metodo rapido e semplice per produrre in vitro grandi quantità di una specifica sequenza di DNA chiamato reazione a catena della polimerasi o PCR. L'amplificazione, che può superare il milione di copie, si realizza in un processo ciclico in tre tempi. I requisiti essenziali della PCR sono:

- sequenza bersaglio di DNA;
- primer forward e reverse ovvero inneschi oligonucleotidici di sintesi di circa 20 nt ciascuno, complementari alle regioni fiancheggianti la sequenza di DNA bersaglio nel campione di DNA;

- una DNA-polimerasi termostabile capace di resistere a 95°C;
- quattro deossiribonucleotidi (A-T-G-C).

La PCR viene svolta attraverso l'applicazione di numerosi cicli di amplificazione e ciascun ciclo prevede tre fasi successive:

- Denaturazione: consiste nella denaturazione termica del campione di DNA mediante riscaldamento a 95°C per circa un minuto;
- Annealing: la temperatura si abbassa lentamente ad un valore ottimale per favorire l'appaiamento degli inneschi;
- Sintesi: la temperatura aumenta fino a circa 72°C per il funzionamento catalitico della DNA-polimerasi Taq (isolata dal batterio *Thermus aquaticus*). Il processo di sintesi ha origine dall'estremità ossidrilica 3' di ciascun innesco.

Questi tre step vengono eseguiti da uno strumento chiamato termociclatore che una volta programmato, permette di passare da una fase all'altra di questo processo ciclico, modificando la temperatura della miscela d'incubazione presente all'interno della microprovetta.

Le sequenze dei primer M13 complementari alle regioni del plasmide fiancheggianti l'insero sono:

PRIMER	SEQUENZA
M13 FORWARD	5'TTG TAA AAC GAC GGC CAG T 3'
M13 REVERSE	5'CAG GAA ACA GCT ATG ACC 3'

La miscela di reazione per la prova con gli M13 per il controllo della lunghezza del frammento (inserto) è costituita da:

REAGENTI	QUANTITA' (μ l)
Primer M13 F (300 ng/μl)	1 μ l
Primer M13 R (300 ng/μl)	1 μ l
Tampone 10x PCR Buffer (privo di MgCl₂)	5 μ l
MgCl₂ (50mM)	1,5 μ l
Deossiribonucleotidi trifosfati (0,25 mM) dNTPs	4 μ l
DNA	(*) ¹
Acqua distillata sterile	37,3 μ l
Platinum® Taq DNA Polymerase (2,5 U/μl)	0,2 μ l
Volume finale	50 μl

(*)¹ Le colonie sono state trasferite nella microprovetta contenete la miscela di reazione in condizioni di sterilità mediante il prelievo diretto della colonia con un puntale sterile.

I parametri che sono stati impostati per la reazione sono stati:

1. 94°C per 2 minuti (attivazione della Taq);
2. 94°C per 1 minuto (denaturazione del DNA);
3. 52°C per 1 minuto (annealing);
4. 72°C per 1 minuto (sintesi);
5. 72°C per 10 minuti.

I passaggi 2, 3 e 4 sono stati ripetuti per 20 cicli.

Il prodotto della PCR è stato caricato su gel di agarosio all' 1,5%, affiancato da un marker per verificare la lunghezza dei prodotti ottenuti. Le colonie che

presentano inserto di lunghezza comprese tra i 500 e 1.000 bp sono state sottoposte a estrazione plasmidica e sequenziate.

2.7. Estrazione plasmidica

Il DNA plasmidico è stato estratto dai cloni ricombinanti di *E. coli* con il metodo della lisi alcalina applicato per preparazioni plasmidiche su piccola scala (miniprep). Questa procedura, basata sulla lisi chimica, prevede l'utilizzo di diverse soluzioni con le quali si provoca la rottura della parete e delle membrane delle cellule batteriche e l'isolamento del DNA plasmidico da quello cromosomico. In questo studio è stato utilizzato il kit *Wizard® Plus Minipreps DNA Purification System (Promega Corporation)*, che consta dei seguenti passaggi:

1. prelevare le colonie selezionate con la PCR e farle crescere in terreno LB a 37°C, 190 rpm per tutta la notte;
2. centrifugare le cellule a 10.000 x g per 5 minuti;
3. eliminare il surnatante e aggiungere al pellet 250 µl di *Cell Resuspension Solution (CRA)* e vortexare;
4. aggiungere 250 µl di *Cell Lysis Solution (CLA)* e mescolare la provetta invertendola, senza vortexare;
5. incubare 5 minuti a temperatura ambiente;

6. aggiungere 10 μ l di *Alkaline Protease Solution* e mescolare invertendo la provetta 4 volte;
7. incubare 5 minuti a temperatura ambiente;
8. aggiungere 350 μ l di *Neutralization Solution* e mescolare immediatamente invertendo la provetta 4 volte;
9. centrifugare a 14.000 x g per 10 minuti a temperatura ambiente;
10. trasferire il surnatante su una *Spin Column* inserita in un *Collection Tube* fornito dal kit;
11. centrifugare la colonna contenente il surnatante alla massima velocità per 1 minuto e scaricare il liquido ottenuto nella *Collection Tube*;
12. aggiungere alla colonna 750 μ l di *Column Wash Solution*, centrifugare alla massima velocità per 1 minuto a temperatura ambiente e scaricare il liquido ottenuto nella *Collection Tube*;
13. aggiungere 250 μ l di *Column Wash Solution*, centrifugare alla massima velocità per 2 minuti a temperatura ambiente e scaricare il liquido ottenuto nella *Collection Tube*;
14. trasferire la colonna su una nuova provetta da 1,5 ml;
15. eluire il DNA plasmidico aggiungendo 100 μ l di *Nuclease-Free Water* alla colonna;
16. centrifugare alla massima velocità per 1 minuto a temperatura ambiente.

La corretta estrazione viene verificata attraverso corsa elettroforetica. Successivamente, i plasmidi purificati vengono quantizzati al Qubit® e sequenziati.

2.8. Sequenziamento

Il sequenziamento dei frammenti di DNA delle colonie di interesse è stato affidato al centro BMR Genomics di Padova (<http://www.bmr-genomics.it>), utilizzando i protocolli di preparazione dei campioni e dei relativi primer presenti sul sito. BMR Genomics effettua il sequenziamento standard con metodo originale di Sanger (Sanger et al., 1977), in cui vengono utilizzati ddNTPs marcati con 4 fluorocromi differenti. Questo consente di far correre i frammenti, ricavati dalla reazione di sintesi, in un'unica corsia, costituita da un gel per elettroforesi introdotto dentro un capillare. Durante l'elettroforesi i frammenti, passando attraverso una specifica finestra, vengono sottoposti ad una luce laser, che ecciterà i diversi fluorocromi presenti sui ddNTPs. Un rilevatore è in grado di captare queste diverse lunghezze d'onda e l'intensità dell'emissione fluorescente dei 4 diversi fluorocromi, identificando il nucleotide presente nella banda. Queste informazioni vengono riportate in un grafico denominato elettroferogramma, in cui i risultati vengono trasformati in picchi di colore diverso in base al nucleotide presente, con aree proporzionali all'intensità di emissione.

2.9. Southern blotting

Il DNA genomico di *C. orientalis*, digerito con l'enzima *Pst*I, è stato prima sottoposto ad una corsa elettroforetica su un gel di agarosio al 2% e successivamente trasferito su una membrana di nylon attraverso i seguenti passaggi.

2.9.1. Blotting

1. Lavare il gel con acqua bidistillata;
2. immergere e agitare il gel per 10 minuti nella soluzione di depurinazione (0,25 M HCl);
3. successivamente trattare il gel con acqua distillata, immergerlo nella soluzione di denaturazione (1,5 M NaCl e 0,5 M NaOH) e porre in agitazione per 20 minuti;
4. lavare ulteriormente il gel con acqua distillata, e poi trattarlo con la soluzione di neutralizzazione (1,5 M NaCl e 0,5 M TrisHCl pH 7.4) e agitare per 20 minuti;
5. contemporaneamente tagliare la membrana di nylon secondo le dimensioni del gel e bagnare la stessa con acqua distillata per 1 minuto;
6. bagnare un foglio di carta assorbente con SSC 10 \times , disporlo su un supporto che fungerà da ponte, posto in una vaschetta, e rimuovere le eventuali bolle formate;

7. bagnare il gel con SSC 10×, disporlo con il lato inferiore sopra il foglio di carta assorbente, e rimuovere le eventuali bolle formate;
8. posizionare la membrana di nylon sul gel, bagnare con SSC 10× e rimuovere le eventuali bolle;
9. bagnare un foglio di carta assorbente con SSC 10×, disporlo sulla membrana, e rimuovere le eventuali bolle formate. Ripetere due volte;
10. disporre dei fogli di carta assorbente, precedentemente tagliati con le dimensioni del gel, impilandoli l'uno sull'altro. Infine porre al di sopra della colonna di carta un peso;
11. avvolgere la vaschetta con della pellicola trasparente in modo da evitare l'evaporazione del tampone e incubare per tutta la notte.

2.9.2. Fissaggio del DNA alla membrana

1. Lavare la membrana con SSC 2×;
2. fissare il DNA per 30 minuti a 120°C in stufa;

2.9.3. Marcatura con digossigenina DIG dell'inserito del clone 59p1

L'inserito contenuto nel clone 59p1 è stato amplificato con i primers M13 e successivamente marcato con digossigenina secondo la tecnica utilizzata per la marcatura del DNA di *C. orientalis* descritta nella sezione 2.5.1 ed utilizzato come sonda. La quantificazione e determinazione dell'efficienza della sonda, l'ibridazione, i lavaggi post-ibridazione e la determinazione immunologica

sono stati effettuati secondo le procedure riportate nei paragrafi: 2.5.2, 2.5.4, 2.5.5 e 2.5.6.

La temperatura d'ibridazione della sonda è stata calcolata considerando il contenuto in GC in percentuale secondo la seguente equazione:

$$T_m = 49,82 + 0,41 \times (\% G + C) - \left(\frac{600}{\text{lunghezza del frammento}} \right)$$

$$T_{\text{ottimale}} = T_m - 20 \text{ } ^\circ\text{C}$$

$$T_{\text{ottimale}} \text{ C159p1} = 49,01 \text{ } ^\circ\text{C}$$

2.10. Dot Blot quantitativo

Il Dot Blot quantitativo è stato utilizzato per valutare la quantità degli elementi C159p1, C17p1 e C11p2 nel genoma di *C. orientalis*, utilizzando il kit *DIG High Prime DNA Labeling and Detection Starter Kit*, il quale impiega digossigenina (DIG), un aptene steroideo, per marcare sonde a DNA per l'ibridazione.

Il metodo permette, inoltre, attraverso un saggio immuno-enzimatico, di sviluppare colore con intensità direttamente proporzionale alla quantità di DNA che si è legato alla sonda marcata.

A tale scopo è stata confrontata l'intensità del segnale di ibridazione dell'amplificato della sonda marcata, ibridizzata con sé stessa, con l'intensità dell'ibridazione della sonda ibridizzata con il DNA di *C. orientalis*, entrambi a concentrazione nota, per risalire poi alla percentuale di DNA ripetuto nel genoma.

Le diluizioni del DNA genomico e delle sequenze inserto sono state riportate in tabella 2.

TABELLA 2

Provetta	Concentrazione DNA genomico	Concentrazione degli inserti
1	100 ng/μl	1 ng/μl
2	50 ng/μl	0,5 ng/μl
3	25 ng/μl	0,25 ng/μl
4	12,5 ng/μl	0,125 ng/μl
5	6,25 ng/μl	0,0625 ng/μl
6	3,125 ng/μl	0,03125 ng/μl
7	1,56 ng/μl	
8	0,78 ng/μl	

Un microlitro delle diluizioni ottenute per il DNA e per le sequenze inserto è stato spottato su una membrana di nitrocellulosa. Successivamente si è proceduto con la denaturazione e la fissazione tramite i seguenti passaggi:

- porre la carta assorbente sul fondo di quattro piastre Petri;
- eliminare gli eccessi del liquido ed evitare la formazione di bolle;
- porre la membrana sulla prima piastra contenente la soluzione SDS al 10% per 3 minuti;
- trasferire la membrana sulla seconda piastra contenente *Denaturing Solution* (NaCl 1,5 M e NaOH 0,5 M) per 5 minuti;
- trasferire la membrana sulla terza piastra contenente la *Neutralizing Solution* (NaCl 1,5 M e Tris-HCl 0,5 M, pH 7,4). Ripetere per due volte;

- trasferire la membrana sulla quarta piastra contenente la soluzione di SSC 2x per 5 minuti;
- asciugare la membrana per 30 minuti a temperatura ambiente;
- fissare il DNA per 30 minuti a 120 °C.

Per le sequenze inserto dei cloni 59p1, 7p1 e 1p2 sono state ottenute le sonde come descritto nella sezione 2.9.3. La quantificazione e determinazione dell'efficienza delle sonde, l'ibridazione, i lavaggi post-ibridazione e la determinazione immunologica sono stati effettuati secondo le procedure riportate nei paragrafi: 2.5.2, 2.5.4, 2.5.5 e 2.5.6.

Anche per C17p1 e C11p2 è stata calcolata la temperatura di ibridazione, secondo la formula riportata nella sezione 2.9.3 ed i valori ottenuti sono i seguenti:

- T ottimale C17p1= 49,33°C;
- T ottimale C11p2= 49,37°C.

2.11. Analisi bioinformatica delle sequenze

2.11.1. Dotlet

Al fine di indagare la presenza di ripetizioni nella sequenza d'interesse, è stato utilizzato il programma Dotlet, (Pagni & Junier, 1999) il quale consente di comparare le sequenze mediante il metodo del plot diagonale, mostrando come risultato del confronto, una rappresentazione grafica della matrice di

allineamento che deve essere interpretata visivamente. Questo programma ha consentito di confrontare le sequenze con sé stesse, allo scopo di individuare l'organizzazione della sequenza ripetitiva del clone 59p1.

2.11.2. Ricerca delle sequenze nelle banche dati

Le sequenze d'interesse 59p1, 7p1 e 1p2 sono state analizzate nelle banche dati.

La ricerca nelle banche dati possiede diverse utilità:

1. verificare la validità di una sequenza dopo la sua lettura, assicurandosi che il sequenziamento non sia frutto di contaminazione;
2. confrontare le sequenze di DNA d'interesse con quelle presenti in banche dati;
3. reperire sequenze di altre specie, appartenenti alla stessa famiglia dell'organismo da cui è stata estratta la sequenza d'interesse, al fine di ampliare l'indagine filogenetica.

Il confronto delle sequenze con quelle presenti in banca dati è stato effettuato attraverso il programma BLAST (*Basic Local Alignment Search Tool*) (Altschul et al., 1990) disponibile alla pagina internet <http://blast.ncbi.nlm.nih.gov/Blast.cgi> del *National Center for Biotechnology Information* (NCBI), gestito dalla *Library of Medicine* (Bethesda USA).

La ricerca è stata svolta attraverso *Nucleotide* BLAST, il quale consente di comparare tra loro diverse sequenze nucleotidiche, ma anche attraverso

BLAST X, il quale consente il confronto delle sequenze nucleotidiche con sequenze proteiche

2.11.3. *Analisi delle sequenze con il programma CENSOR*

CENSOR (Jurka et al., 2005) è un software che consente di confrontare le sequenze inserite con una collezione di riferimento di sequenze di elementi ripetuti ricercando fra queste porzioni omologhe e generando infine un rapporto su tutte le corrispondenze trovate. Tale strumento è stato utilizzato per evidenziare l'eventuale correlazione tra le sequenze 7p1 e 1p2 con elementi trasponibili.

2.11.4. *Analisi delle sequenze con il database CDD*

Il database CDD (*Conserved Domain Database*) è una fonte di annotazione di proteine per l'identificazione rapida di domini conservati in sequenze proteiche attraverso RPS-BLAST. CDD evidenzia la relazione tra sequenza, struttura e funzione. In questo studio è stato utilizzato per individuare i domini strutturali che caratterizzano le sequenze dei cloni 1p2 e 7p1.

2.11.5. *Analisi filogenetiche*

Gli alberi filogenetici sono stati costruiti utilizzando l'allineamento delle sequenze ottenute con ClustalW (Larkin et al., 2007) utilizzando il metodo di Inferenza Bayesiana (*Bayesian Inference*) condotto con il programma Mr. Bayes (Huelsenbeck et al., 2002). Gli alberi sono stati inoltre radicati attraverso

il metodo del mid-point rooting in quanto non si disponeva di un outgroup ottimale al fine di orientare gli alberi.

Nel caso dell'analisi condotta per il clone 1p2 sono state utilizzate sequenze amminoacidiche e il modello evolutivo è stato selezionato attraverso il programma Mr. Bayes ed è risultato il modello WAG. Il numero di generazioni è stato impostato a 1000000 con una frequenza di campionamento ogni 100; il burn-in, pari a 1000, rappresenta un quarto delle generazioni campionate.

Per il clone 59p1 l'analisi è stata condotta su sequenze nucleotidiche e la specificazione del modello evolutivo è stata effettuata attraverso l' Akaike Information Criterion (AIC), utilizzando il programma Model Test 3.8 (Posada e Crandall, 1998), tra una serie di possibili modelli calcolati sulla base dell'allineamento multiplo utilizzato. Il modello evolutivo selezionato è TPM3+ G (Rodríguez et al., 1990): la frequenza delle basi è uguale per tutte e il parametro di distribuzione Gamma è 2,4150. La matrice di sostituzione propone A- C =0.5074; A- G =1.9565; A- T =1.000; C- G =0.5074; C- T =1.9565; G- T =1.0000, e la proporzione di siti invariati è = 0.000. Il numero di generazioni è stato impostato a 4000000 con una frequenza di campionamento ogni 100; il burn-in, pari a 10000, rappresenta un quarto delle generazioni campionate.

Per tutte le analisi condotte per valutare l'affidabilità dei nodi vengono utilizzati i valori di una probabilità a posteriori. Un nodo è considerato affidabile se il valore di probabilità a posteriori raggiunge o supera 95.

Capitolo terzo

RISULTATI

3.1. Valutazione generale delle sequenze ottenute

Il DNA estratto dal tessuto muscolare di *C. orientalis*, mediante alcool isoamilico e cloroformio (metodo CIA), è stato quantizzato al Qubit™ che ha rivelato una concentrazione nel campione pari a 191 ng/μl.

La libreria genomica costituita da 86 colonie è stata ottenuta in seguito al clonaggio dei frammenti derivanti dalla digestione del DNA di *C. orientalis* con l'endonucleasi di restrizione *Pst*I. Le colonie di interesse sono state trasferite su membrane di nitrocellulosa e successivamente ibridate con il DNA genomico di *C. orientalis* marcato con digossigenina (FIGURA 10). Le colonie mostranti un segnale più intenso, indice della presenza di elementi ripetuti, sono state amplificate mediante PCR con i primers M13 per verificare la lunghezza dell'insero; successivamente si è proceduto con l'estrazione plasmidica e il sequenziamento.

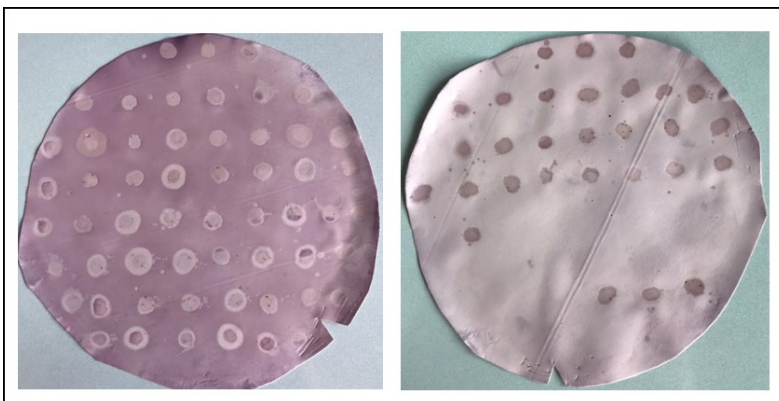


FIGURA 10 Ibridazione delle colonie della libreria genomica usando come sonda il genoma di *C. orientalis*.

Sono stati sequenziati 16 cloni che presentano lunghezza dell'inserto molto diversa. Le analisi condotte sulle banche dati hanno portato a identificare delle sequenze ripetitive in tandem e altre intersperse. In altri casi la ricerca non ha prodotto risultati rilevanti (TABELLA 3).

TABELLA 3 Cloni sequenziati

NOME CLONE	COMPLETEZZA SEQUENZA	NCBI blastn	NCBI blastx	CENSOR
<i>Cl1_p2F_R</i>	COMPLETA 1035 bp			LTR Gypsy
<i>Cl2_p2F</i>	COMPLETA 277 bp			
<i>Cl10_p1F</i>	COMPLETA 881 bp			LTR copia22
<i>Cl7_p1F</i>	COMPLETA 847 bp			Non-LTR
<i>Cl18_p1F</i>	COMPLETA 178 bp			
<i>Cl25_p1F</i>	COMPLETA 706 bp	microsat		
<i>Cl31_p1F</i>	COMPLETA 971 bp	satDNA		ATSAT2
<i>Cl37_p1F</i>	INCOMPLETA			
<i>Cl45_p1F_R</i>	COMPLETA 961 bp			LTR DIRS
<i>Cl51_p1F</i>	COMPLETA 698 bp	microsat		ERV
<i>Cl52_p1F</i>	COMPLETA 485 bp			ERV
<i>Cl54_p1F</i>	INCOMPLETA			LTR, nonLTR, DNA transp
<i>Cl56_p1F</i>	COMPLETA 331 bp			LTR Gypsy
<i>Cl57_p1F</i>	COMPLETA 338 bp		CGG triplet repeat- binding 1	
<i>Cl59_p1F</i>	COMPLETA 668 bp	DNAst2		ATSAT2
<i>Cl60_p1F</i>	COMPLETA 379 bp		Zinger finger protein	

Data l'abbondanza di risultati ottenuti, per questo studio si è scelto di soffermarsi principalmente su tre cloni: *cl1p2* che presenta una sequenza inserto appartenente ad un LTR della famiglia *Ty3/Gypsy*, *cl7p1* che presenta

una sequenza inserto appartenente ad un retrotrasposone della famiglia DIRS ed infine cl59p1 che presenta una sequenza inserto appartenente ad una sequenza ripetuta in tandem.

3.1.1. Risultati della sequenza inserto del clone 1p2

La sequenza dell'inserto presenta una lunghezza di 1035 bp e ha un contenuto in percentuale di nucleotidi GC del 49,1%.

L'analisi della s*equenza è stata condotta innanzitutto su BlastX, dove si è rilevata una percentuale di identità di sequenza bassa, inferiore al 55%, con proteine non caratterizzate di *Xenopus tropicalis* e con una proteina *pol* derivata da retrovirus presente in elementi trasponibili in diversi organismi.

L'analisi su Censor ha evidenziato la similarità della porzione iniziale (585 bp) del clone di interesse con la sequenza di un elemento LTR della famiglia *Ty3/Gypsy* presente in un rettile (FIGURA 11).

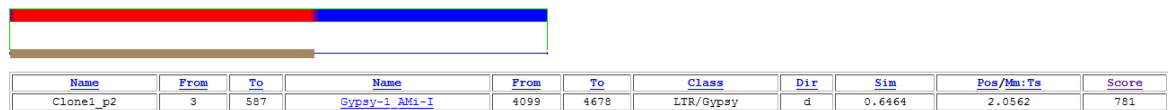


FIGURA 11 Risultato dell'analisi sulla sequenza cl1p2 su Censor.

Continuando l'analisi con ORF finder, al fine di individuare le ORF presenti nella sequenza di interesse, è stato possibile individuare la ORF3 (FIGURA 12) che presenta la seguente sequenza:

>ORF3

KAGLTIKATKQCIGQSQVVYLGHVLVGGGQVQPLRSKIQAIMDWVAPTQTQVRAFLGLTGYYRKFV
RGYGSIVTPLNDLTSKKMPKKVQWTASCQKAFDDLKQAMCTAPVLQSPDYTKKFIVQTDASELGVG
AVLSQLNKDGDQDPVAFISRRLTPREKRWSAIEREAFAVVWALKKLRPYLFGTHFIVQTDHKPLLW
LKQMKGENPKLLRWSISLQGM DYTVEHRPGTQHANA DGLSRYFHLEHEDLSGQG

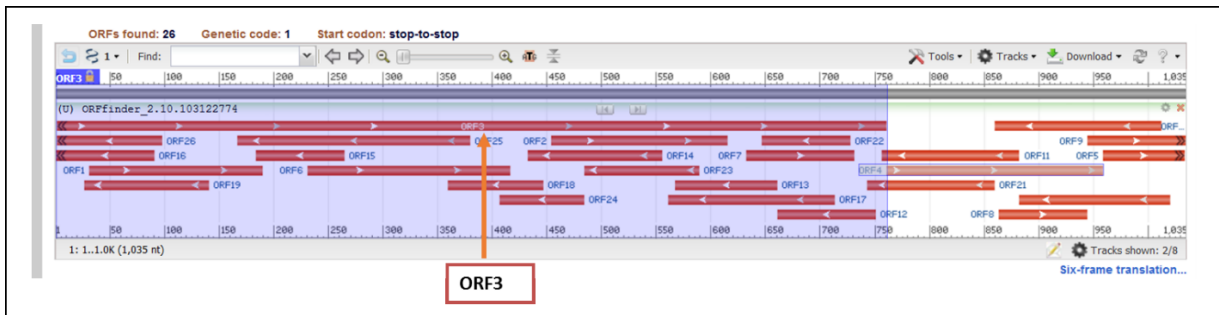


FIGURA 12 Risultato dell'analisi condotta con ORF finder sulla sequenza inserito del clone *cl1p2*. La sequenza indicata è l'ORF3.

Analizzando la sequenza ORF3 su BlastX si evince che presenta similarità con la sequenza per la proteina RNAsi-H presente in elementi *Ty3/Gypsy*. Inoltre questa sequenza ha il sito catalitico DEDD (lettere evidenziate in giallo nella sequenza sopra) conservato e questo porterebbe a ipotizzare che sia ancora attivo.

A partire dalla sequenza dell'inserito del clone 1p2 è stato possibile, inoltre, effettuare un'analisi sul trascrittoma di *C. orientalis*. Il risultato di questa analisi ha permesso di identificare due sequenze del trascrittoma, che come si nota dall'allineamento riportato sotto, presentano un tratto conservato con la sequenza d'interesse e sono di lunghezza maggiore.

Allineamento del clone lp2 con i primi due risultati ottenuti dall'analisi condotta sul trascrittoma

TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TTAAACCATAAGGGATTGTTTGACCTGGTGCAAGGTGACCACTAGAACTAGTGGGCAGCA	60
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TATACCAGGCCAACATCTCACATTGGTGAGCAGCGTGTGGGATTCCAGCTTGCTATTGTT	120
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TTACCAAATAATCTCAAAGTTTATATCAACTGAAACCACTATAGGTTCTCGAGGAACAT	180
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	ATAATACTAGCTAGATATACAAGTTCTACTGAACCTAGAGTTGTGGGCAGAGGGACAAGC	240
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	ACACCCTCCTCCACAACTCAAGACAGGGTCAGTAGGACACCACTAGTGTAGGTAGTGCA	300
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TACACCACACCTGCATTAGACATGTCTTTAAAAGGTGTGAGCCAAGCCTCTACACCCTCT	360
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GCCTATGAGGACCTGCTCTTTAATGAGCTGAGGGCCCTCACCAAGGCAAGAGGAATAAAG	420
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	ACAGAGAAACATCCTACCCAGCAGCAATTGATAGCAATTCTGCTAGAGGATGATCTCATG	480
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	AAGGCCTTGGGCCCTCAAGGGGATCATGGGTGGGACCCAGCACTCAGATAAAGGCTCCC	540
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	AATGGTAATAACCAGGGGAGTCTGACTCTGAGGATGCCCCAGAGAACCAACAGGATTCA	600
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	AATCTGGGCCATCATCCTAGATCTAGGGGTGAGGAGAGGAGAAGCCCTAACAGGCCCTTC	660
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TTTAATCTCCCTGGGGAAGGTGATGGGGCACTTGGGCGCAGTAGAACACCTTCAGTGGAT	720
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TCTAACTGCTCCAAGGCATCAAAGAAGGCCCTCTGCCTCTCTTACAGATAAGAGAGAATCC	780
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0

TRINITY_DN108633_c1_g1_i4	TTCATGAGAGGAAGTGCAGACTGCTGGAGTTTGAGGAAGCAAACCTCAGACTCAAGAGA	840
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GAGCAGTTAGCACTAGACAGGGAATCCCTGGAGCTCCAGAGGGCTTCAAAGAAATTGGGG	900
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TTAGACCCCTATGGTGGCAGCAACCCTATCCCTAGAAGTAGAGATAGGGATAGTTATGAG	960
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	CATAAGAACTTGAGTAAGATTGTGCCCCCTATCAGAGTGGGGATGACATCCACAAGTGG	1020
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	ATAACTGCTATGGAGAGGGCTGTAAAGGTACAAGGGTCCCTCCCAGGCAGTGGGCAGCA	1080
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	ATCTTATGGCTCTCCTTTGCAGACAAGGTAGAGACAAGCTCCTCACTGTTAGAGAGGAG	1140
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GATGCTGATGACTATATGGTCTGAAGAGTGCTCTTATTGATGGGTTTGGCTTAACCACT	1200
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GAACAGTATAGGTTAAAGTTCAGAGACACCCAAAAGGCACCAAATCAGACTGGACAGAC	1260
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TATGTAGATTGTGCAACCAAGGCTCTGCAGGGTTGGTTGATTGGCAGTAAGGTAACCACA	1320
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TTTTTAGGATTGTATAATCTAGTCAATGAGGAACAGATCCTTAATAACTGTGAGAACCAG	1380
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	CAGTTGCATCAACACCTGGTTGACTCAGATCTGACCTCTCCCCAAGAATTGGGAAAGAAG	1440
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GCAGACAAGTGGTTCAGGACCAGGGTGAAGCAGAAAACCCAGAACCAAGGGGGCAACTCT	1500
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	CAGAAAGCAGATGGTAAGGCTCTGACAAGGGGGTGAAGAGTCTAAAGACCAGAAGTCT	1560
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TCCAAGGGCCACAAAATCTGGGGGGAGTCACAATCTAACAAGAAAACCAAGACAAG	1620
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	AAGAAGTCCTATTGTAAGTTTGTAAAGATAAAGGGCCATTGGACTTCTGATTGCACCAAG	1680

TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TCTAAGGACTCAACCACCACTACTACTGCCCTAAGGCTCTACTCCCCCTAGCACAGTT	1740
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	ACTAGTGGAGGTGGTACACAGCAGAAGGGTGTAGTGGGTTTGACCATGGGAAGTGGGGT	1800
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GACTCTGGCCTAGTCAGAGTAGCCACTCAGGCATTGCTAGTCTGTGATGGAGACATAGAT	1860
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	ATGTCATTCTAGTGTAGTCCCAGTAATCTGGGCAAATACAGACAGCAGCCTGTAATCAAT	1920
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GGGAGAGAGACCTTAGCATTTCAGGGACACAGGTGCTAGTGTCCACCATGGTGGCAGAGGAC	1980
*_		
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	CTTGTGCCTCCTGAGCAGATCATCTGGGTGAAATGTACCAGGTGACCAACGCAAGCAAT	2040
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GTCCAGACCAACCATAGTCTGGCAGTTGTGAACTTTCCTGGGGAGGGGTAAGTGGCCTA	2100
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	AAGAAGGTTGTTGTAAGTGCAGACCTCCCTGTAGAATGTTTGTGGGCAATGATCTTGAA	2160
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GCCTCAGCCTGGTCTGAGGTGGAGTTAAGGACTCATGCAGAGATGCTTGGCATACTGCA	2220
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GAGGTTTTTGTGTGACCAGGTCAACAAGCAAAGATGCAAAACACTGAACCAGCACTGGAT	2280
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	CCTGGAACAATGGACCAGTTGCAATCCAAGAGAAGGAGCAGCAGGGTAAAGGCACAGTCT	2340
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TCCCCTCTGTCTCAGGCAGATGACCCTCCTATGGAGGAGGAGAATCTAAATCCTGAAGAA	2400
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GGAAGTGCACCAGAGGAGCTTGAATCTCCCAGATTGAGTTCTTGGGTGCAGGGGGACCT	2460
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TCTAGGGAGGACCTGAGCAAAGAGCAGCAGTGCTGCCAACTCTTGAACCCCTCAGACAG	2520
TRINITY_DN108577_c4_g2_i3	-----	0

Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	CTAGCTGCTGCTGAGGCTGCTGGGAGTGCCAGTGGGTCCCATAGAGTTTATTGGGATGGG	2580
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GACCTCCTGTATTCTGAGGGAAGAGACCCTGCCCTGGAGCCACCAGGAGGCTGGTTATT	2640
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	CCACAGAAATTTAGGGAGTTCCTCTTGACCCTGGCACATGATATCCCATTGGCAGGACAT	2700
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TTAGGCCAGAATAAGACTTGGGACAGGATTGTTCCACACTTTCCTGGCCTCACATGTCA	2760
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GAGGATGTGAGAGACTTTTGTAAAGTCTTGTGTAACCTGTCAGGCCAGTGGCAAGACAGGA	2820
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	AACACACCAAAGGCCCCCTGATACCACTACCTGTGGTGGGGTTCCCTTTGAAAGGGTA	2880
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GGGATTGACATTGTGGTCCCCTTGACCCTCAAACAGCATCTGGGCACAGATTCATACTA	2940
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GTGGTGGTGGATCACGCCACTAGGTATCCAGAGGCAATTCCTTAGGACCACTACAGCA	3000
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	CCTGTTGTGCAAGGCCCTCATTGGAATATTTCCAGAGTGGGCTCCCTAAGGAGGTG	3060
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	ATATCTGACAGAGGTGCAAACTTTATGTCAAGTTATCTCAAGGCTATGTGAAAGAGTGT	3120
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GGGGTGACCTACAAGTTCCTACACCATACCACCCACAATGCAATGGGCTTGTGAAAGG	3180
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TTCAACAAGACCCTAAAGGGCATGATCAAGGGGCTCCAGAAAAGCTTAGGCAGAAAATGG	3240
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GACTTGCTCTTGCCCTGCCTTCTATTTGCATATAGGAGGTTCCCTCAGAAGGGGTGGC	3300
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TACAGCCCTTTTGAACCTTCTGTTTGGGCACCCTGTAAGAGGTCCCCTGTGTCTGGTGAAG	3360
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GAAGGTTGGGAGTTACCTCTCAAACCCCTAAGCAAGACATGTGGGCTATGTACTTGGT	3420

TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	CTCCGCTCCAGGATGGCTGAGTATATGAAGAAGGCAAATAAAAACCTGGAGGCAAGTCAG	3480
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GAATTACAAAAGGCCTGGCATGACCAAAGGCTGAACTGGTGGTTTACCATCCAGGGCAG	3540
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	AAAGTGTGGGTGTTGGAGCCTGTGGCCCCAGCAGCCCTCCAAGACAAGTGGAGTGGCCCC	3600
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	CATACAATTGTTGAGCAGAAGGGAGAGGTTACCTACCTGGTGGACCTTGGAACTCCCAGG	3660
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	AGCCCCCTAAGGGTGTTCATGTCAACAGACTCAAACCATATGTGGATAGGGCGGACATG	3720
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GGCTTGCTCATGGCCACTGATGAGGGACAGGAGGAGAGTACCCCTCTCCCTGACCTG	3780
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	CTCTCACAGGACCCTAAGGATGGTTCAGTAGAAGGTGTTGCTTTGGCAAGCAGCCCTACT	3840
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GACATCCAACAGACTGAGATCAGGCAAGTCTGAACCAGTCTCCACATATTCAGTCTG	3900
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	ACCCCTAGTCAGACTTCTGGTGCACCCATGAGGTGGACACTGGAGACAGTCTCCCTGTA	3960
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	AAGAGTAAATCTACAGGCAAGCAGATCATGTAAGAGACTGTATCAGAGCTGAAGTCTGTG	4020
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	AAAATGCTGGAATTGGGTGTGATTGAGCCTTCAGCTAGCCCTGGGCTAGTCCAGTAGTC	4080
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TTGGTCCCAAACCTCATACCCAGGGTGGTAAGAAAGAGATGAGGTTTTGTGTTGACTAC	4140
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	AGAGGCCTCAATGCTGTTACTAGAACAGATGCATATCCTATTCCTAGGGCAGATGAGCTG	4200
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	ATTGATAAGCTGGCCTCAGCCACTTACCTCAGTACATTTGACCTGACAGCAGGGTATTGG	4260

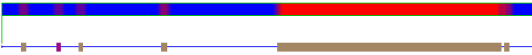
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	CAAATCAGACTAGCTGAGGAAGCCAAGCCAAGACAGCATTCTCAACCCCTGGTGGGCAT	4320
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	TATCAGTTTGTGGTAATGCCCTTTGGTCTGAAGAATGCACCTGCCACCTTCCAAGGTTG	4380
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GTGAATCAAGTCTGGCTGGCTTGAAGACTTTTGTGCAGCATATCTAGATGATATTGCT	4440
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	-----	0
TRINITY_DN108633_c1_g1_i4	GTCTTTAGCTCCAGCTGGCAGGATCACCTGATCCACCTGAGGAGAGTGTGGAGGCCCTG	4500
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	--GAAAGCAGGCCTCACTATCAAGGCCACTAAGTGCCAGATAGGGCAGAGCCAGGTTGTG	58
TRINITY_DN108633_c1_g1_i4	CAGAAAGCAGGCCTCACTATCAAGGCCACTAAGTGCCAGATAGGGCAGAGCCAGGTTGTG	4560
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	TACCTGGGCCACCTTGTGGGTGGAGGCCAGGTTCAACCACTAAGGAGCAAAATCCAGGCC	118
TRINITY_DN108633_c1_g1_i4	TACCTGGGCCACCTTGTAGGTGGAGGCCAGGTTCAACCACTAAGGAGCAAAATCCAGGCC	4620
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	ATAATGGACTGGGTGCTCCAACCTACCCAGACTCAGGTGAGAGCCTTTTGTAGGCCCTGACT	178
TRINITY_DN108633_c1_g1_i4	ATAATGGACTGGGTGCTCCAACCTACCCAGACTCAGGTGAGAGCCTTTTGTAGGCCCTGACT	4680
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	GGGTATTATAGGAAGTTTGTGTAGAGGGTATGGATCCATTGTGACACCCCTGAATGACCTC	238
TRINITY_DN108633_c1_g1_i4	GGGTATTATAGGAAGTTTGTGTAGAGGGTATGGATCCATTGTGACACCCCTGAATGACCCC	4740
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	ACTTCTAAGAAAATGCCAAAAAGGTTTCAGTGGACAGCTAGCTGTCAAAGGCATTTGAT	298
TRINITY_DN108633_c1_g1_i4	ACTTCAAAGAAAATGCCAAAGAAGGTTTCAGTGGACAGCTAGCTGTCAAAGGCATTTGAT	4800
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	GACCTCAAGCAGGCTATGTGCACTGCTCCTGTTCTGCAAAGTCTGACTACACCAAGAAG	358
TRINITY_DN108633_c1_g1_i4	GACCTCAAGCAGGCTATGTGCACTGCTCCTGTTCTGCAAAGTCTGACTACACCAAGAAG	4860
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	TTCATTGTTCAAACAGATGCTTCAGAGCTGGGTGTGGGAGCTGTACTTTCACAATTGAAT	418
TRINITY_DN108633_c1_g1_i4	TTCATTGTTCAAACAGATGCTTCAGAGCTGGGTGTGGAAGCTGTACTTTCACAATTGAAT	4920
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	AAGGATGGTCAGGATCAGCCAGTAGCTTTCATTAGCAGAAGGCTGACCCCCAGAGAGAAA	478
TRINITY_DN108633_c1_g1_i4	AAGGATGGTCAGGATCAGCCAGTAGCTTTCATTAGCAGAAGGCTGACCCCCAGAGAGAAA	4980
TRINITY_DN108577_c4_g2_i3	-----	0
Clone1_p2	CGTTGGTCAGCTATAGAGAGGGAGGCCTTTGCTGTGGTTTGGGCCCTGAAGAAGTTGAGA	538
TRINITY_DN108633_c1_g1_i4	CGTTGGTCAGCTATAGAGAGGGAGGCCTTTGCTGTGGTTTGGGCCCTGAAGAAGTTGAGA	5040
TRINITY_DN108577_c4_g2_i3	-----TTCAGACAGACCATAAGCCCCCTCTATGGCTA	32
Clone1_p2	CCATACCTGTTTGGGACTCACTTCATAGTTTCAGACAGACCACAAGCCCCCTCTATGGCTA	598
TRINITY_DN108633_c1_g1_i4	CCATACCTGTTTGGGACTCACTTCATAGTTTCAGACAGACCACAAGCCCCCTCTATGGCTA	5100

TRINITY_DN108577_c4_g2_i3	AAGCAGATGGAGGGTGAAAACCCCAAACCTTTGAGGTGGTCCATTTCCCTACAGGGGATG	92
Clone1_p2	AAGCAAATGAAGGGTGAGAACCCCAAACCTTTGAGGTGGTCCATTTCCCTACAGGGGATG	658

TRINITY_DN108633_c1_g1_i4	AAGCAAATGAAGGTTGAGAACCCCAAACCTGTTGAGGTGGTCCATTTCCCTCCAGGGGATG ***** ** * * * * * *****	5160
TRINITY_DN108577_c4_g2_i3	GACTATACAGTGGAACATAGACCTGGGACCCAACATGCCAATGCAGATGGACTCTCTAGG	152
Clone1_p2	GACTATACAGTGGAACATAGACCTGGGACCCAACATGCCAATGCAGATGGACTCTCTAGG	718
TRINITY_DN108633_c1_g1_i4	GACTATACAGTGGAACATAGACCTGGGACCCAACATGCCAATGCAGATGGACTCTCTAGG *****	5220
TRINITY_DN108577_c4_g2_i3	TATTTCCACTTAGAACATGAAGACTCAACTGGTCAAGGTTAGTTTAAATGTCCTGCGTTT	212
Clone1_p2	TATTTCCACTTAGAACATGAAGACTTATCTGGTCAAGGTTAGTTTAAATGTCCTGCGTTT	778
TRINITY_DN108633_c1_g1_i4	TATTTCCACTTAGAACATGAAGACTCATCTGGTCAAGGTTAGTTTAAATGTCCTGCGTTT ***** * *****	5280
TRINITY_DN108577_c4_g2_i3	GGGGTG----GTTTGTGTAGGAGAGTAGATCTGCCTATGCTGTTGCCCCACACT-TTTT	267
Clone1_p2	GGGGGGG---GTTTGTGTAGGAGAGTAGATCTGCCTATGCTGTTGCCCCACACT-TTTT	834
TRINITY_DN108633_c1_g1_i4	GGGGGGGGGGGTTGTGTAGGAGAGTAGATCTGCCTATGCTGTTGCCCCACACTTTTTT **** * *****	5340
TRINITY_DN108577_c4_g2_i3	TGATGCTGAACACTGGTTTTTTGACCTTAGTCAGGCTGGCCCTGCTAATCAGGCCCCAGC	327
Clone1_p2	TGATGCTGAACACTGGTTTTTTGACCTTAGTCAGGCTGGCCCTGCTAATCAGGCCCCAGC	894
TRINITY_DN108633_c1_g1_i4	TGATGCTGAACACTGGTTTTTTGACCTTAGTCAGGCTGGCCCTGCTAATCAGGCCCCAGC *****	5400
TRINITY_DN108577_c4_g2_i3	CTCTCTGTTTCCCTCTAAACGAATTGGTGTGCACCTTAGGTGCACCTAGTGATTGGATACT	387
Clone1_p2	CTCTCTGTTTCCCTCTAAACAAATTGGAGTGCACCTTAGGTGCACCTAGTGATTGGATACT	954
TRINITY_DN108633_c1_g1_i4	CTCTCTGTTTCCCTCTAAACACATTGGTGTGCACACTAGGTGCACCTAGTGATTGGATACT ***** * * * * * *****	5460
TRINITY_DN108577_c4_g2_i3	GTGG-TACCCTGTGAGTCCCTAGTATATGGTACCTTAGGTACCCAGGGCATTGGTACTTC	446
Clone1_p2	GTAAGTGCCCTGTGAGTCCCCAGTATATGGTACACCAGGTACTAAGGGCATTGGTACTTT	1014
TRINITY_DN108633_c1_g1_i4	GTAAGT----- ** *	5466
TRINITY_DN108577_c4_g2_i3	ACAGGCCTACCAGGGGCTGCAGCAGCAATTGTGCCACCCCTGGGAAGCATTCTAAATCTG	506
Clone1_p2	ACAGGCCTACCAGGGGCTGCA-----	1035
TRINITY_DN108633_c1_g1_i4	-----	5466
TRINITY_DN108577_c4_g2_i3	ACACCACAGGCATGCCATGGCAGCCTGGGTGAGGTGGTGTAACTACCCCTCACTGAGATA	566
Clone1_p2	-----	1035
TRINITY_DN108633_c1_g1_i4	-----	5466
TRINITY_DN108577_c4_g2_i3	TAGG 570	
Clone1_p2	---- 1035	
TRINITY_DN108633_c1_g1_i4	---- 5466	

In particolare la sequenza TRINITY_DN108633_c1_g1_i4 presenta una lunghezza di 5466 bp ed è stata utilizzata per effettuare ulteriori analisi.

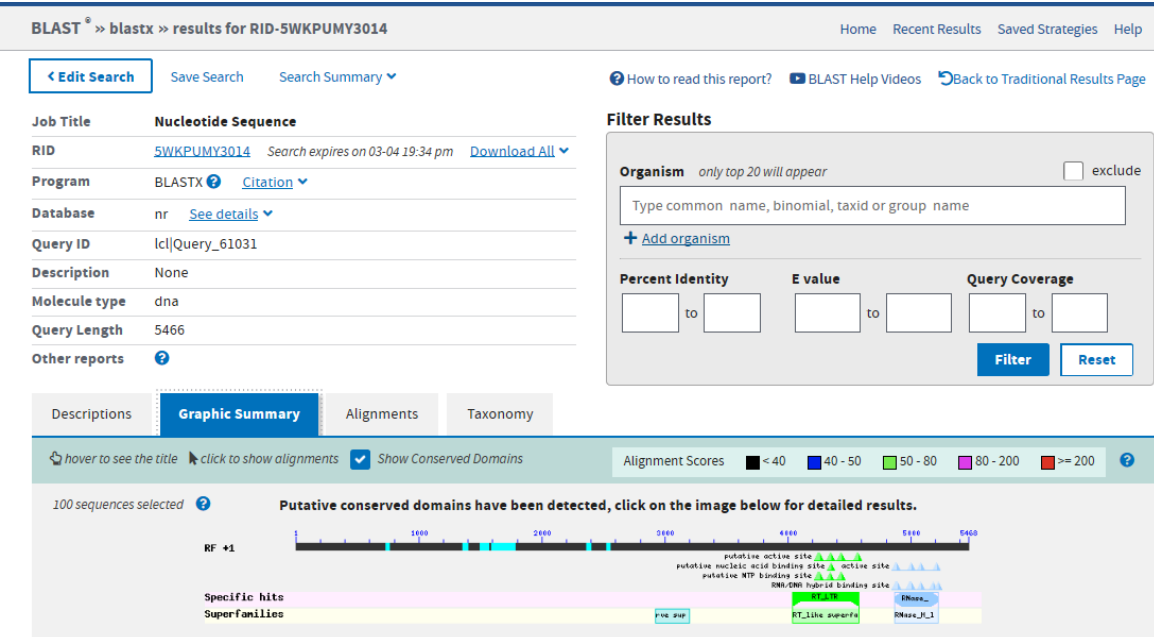
L'analisi su Censor ha permesso di individuare al suo interno cinque regioni contenenti TE diversi, di cui quella più lunga pari a 2235 bp trova similarità con elementi LTR/Gypsy (*FIGURA*).



Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Ms:Ts	Score
TRINITY_DN108633_c1_g1_14	204	252	CINFUL1_ZM_I	6943	6990	LTR/Gypsy	d	0.8163	3.5000	233
TRINITY_DN108633_c1_g1_14	572	609	Mariner-S_DF	1147	1186	DNA/Mariner	c	0.8462	1.6667	237
TRINITY_DN108633_c1_g1_14	796	838	Copia-14_LMi-I	834	874	LTR/Copia	d	0.8333	1.0000	214
TRINITY_DN108633_c1_g1_14	1641	1696	Copia22-ZM_I	3431	3485	LTR/Copia	c	0.7500	1.5000	209
TRINITY_DN108633_c1_g1_14	2840	5139	Gypsy-18_CPB-I	2630	4911	LTR/Gypsy	d	0.6263	2.0241	2235
TRINITY_DN108633_c1_g1_14	5167	5221	Gypsy-12-I_XI	4298	4352	LTR/Gypsy	d	0.7636	2.1667	252

FIGURA 13 Analisi su Censor della sequenza ottenuta dal trascrittoma di *C. orientalis*.

Continuando l'analisi della sequenza del trascrittoma su BlastX è stato evidenziato che sulla seconda parte della sequenza risiedono tre domini: *rve_sup* che è un sito che promuove l'attività dell'integrasi; *RT_LTR* che è il sito per la trascrizione della retrotrascrittasi; ed infine *RNasi-H* è una endonucleasi coinvolta nel clivaggio del RNA ed è lo stesso sito che è stato caratterizzato nella sequenza del clone 1p2 (FIGURA 1414).



BLAST® » blastx » results for RID-SWKPUMY3014

Home Recent Results Saved Strategies Help

[← Edit Search](#) Save Search Search Summary ▾

How to read this report? BLAST Help Videos Back to Traditional Results Page

Job Title: Nucleotide Sequence

RID: SWKPUMY3014 Search expires on 03-04 19:34 pm [Download All](#) ▾

Program: BLASTX Citation ▾

Database: nr See details ▾

Query ID: lc|Query_61031

Description: None

Molecule type: dna

Query Length: 5466

Other reports ?

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity E value Query Coverage

Filter Reset

Descriptions **Graphic Summary** Alignments Taxonomy

hover to see the title click to show alignments Show Conserved Domains

Alignment Scores < 40 40 - 50 50 - 80 80 - 200 >= 200 ?

100 sequences selected Putative conserved domains have been detected, click on the image below for detailed results.

RF #1

Specific hits Superfamilies

[rve_sup](#) [RT_LTR](#) [RNase_H1](#)

FIGURA 14 Analisi su BlastX della sequenza ottenuta dal trascrittoma di *C. orientalis*.

L'analisi condotta con ORF finder sulla sequenza *TRINITY_DN108633_c1_g1_i4* ha mostrato la presenza della ORF1 completamente sovrapposta alla sequenza in analisi (FIGURA 5).

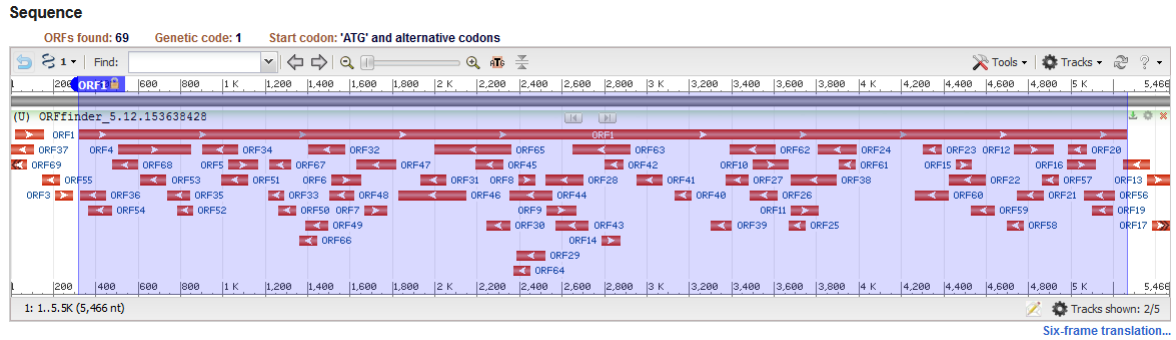


FIGURA 15 Analisi tramite ORF finder della sequenza ottenuta dal trascrittoma di *C. orientalis*.

Tramite il database CDD è stato possibile evidenziare in questa sequenza ancora una volta la presenza di domini tipici degli elementi LTR della famiglia *Ty3/Gypsy* (

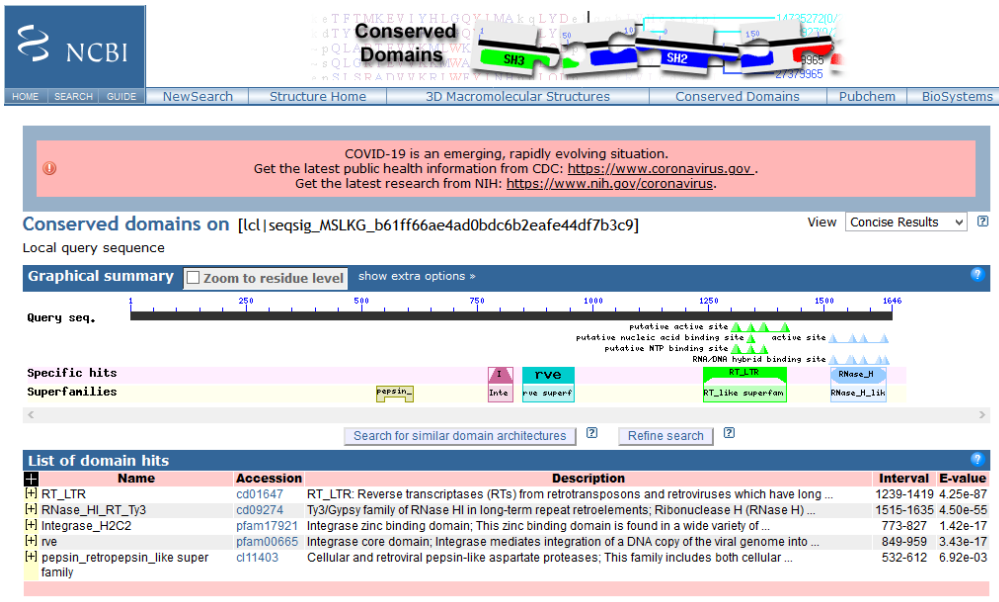


FIGURA 26).

FIGURA 26 Risultato dell'analisi della sequenza ORF1 condotta sul database CDD.

La sequenza della ORF1 è stata analizzata anche su BlastP e se ne è ricavato un allineamento con proteine non caratterizzate di diversi organismi tra cui *X. tropicalis*, *X. Laevis* e *A. carolinens*. Tutte le sequenze sono state scaricate in formato fasta ed è stato eseguito un allineamento con MAFFT (gap open 2.5 e gap extension 0.5). Con quest'ultimo è stato costruito un albero filogenetico utilizzando come metodo l'inferenza bayesiana (*FIGURA*).

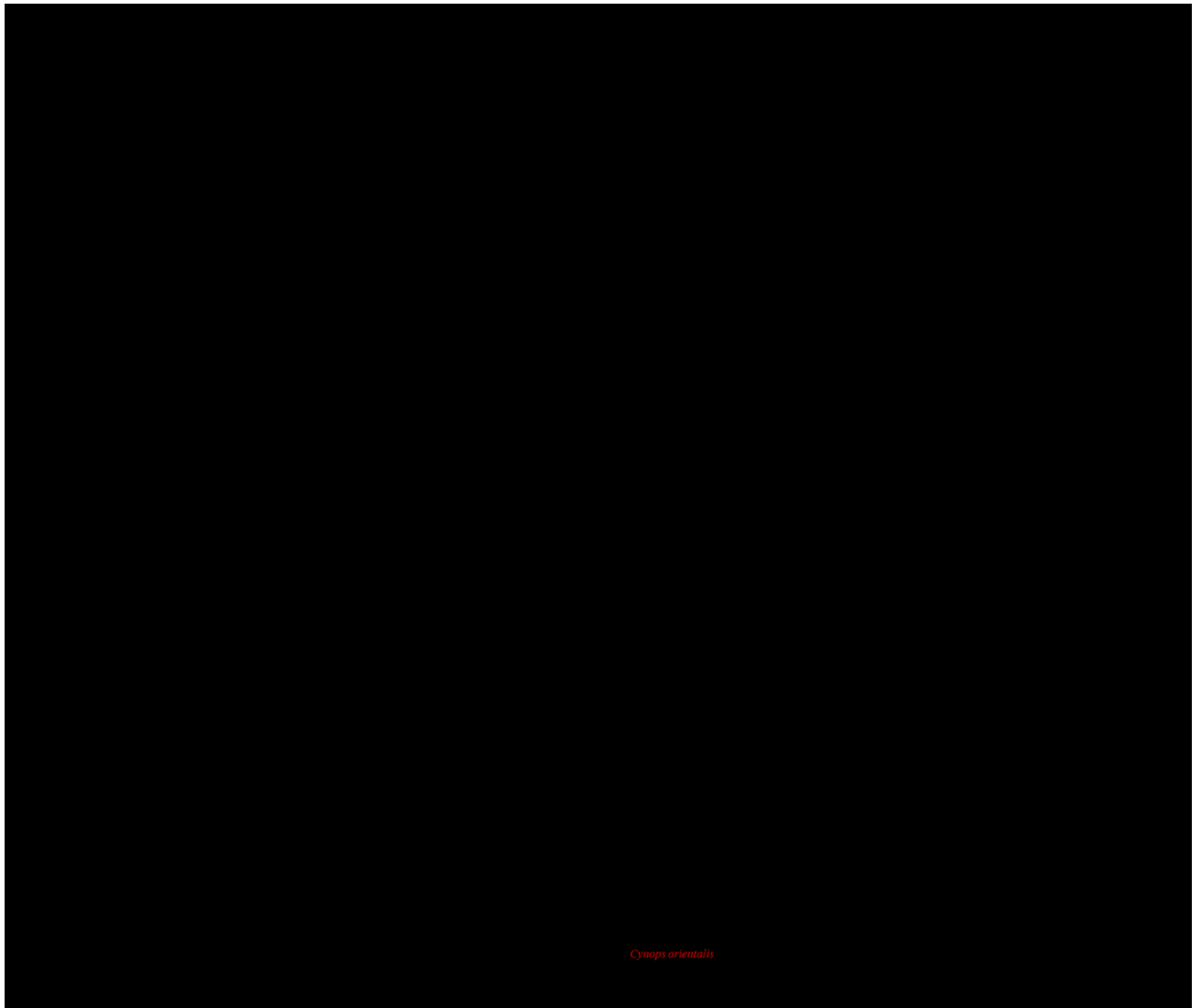


FIGURA 17 Albero filogenetico delle sequenze ottenute su BlastX dall'analisi della sequenza ORF1.

L'analisi filogenetica (FIGURA 17) evidenzia due raggruppamenti: il primo contiene sequenze appartenenti ad echinodermi, teleostei, celacanti, anfibi e rettili; il secondo oltre a sequenze appartenenti ad echinodermi contiene sequenze di rettili e anfibi. La sequenza di *C. orientalis* si colloca in questo secondo gruppo in posizione esterna alle sequenze di *X. tropicalis*.

È stato infine valutato quanto l'elemento analizzato sia presente nel genoma di *C. orientalis* attraverso un dot blot quantitativo, utilizzando concentrazioni scalari sia della sequenza cl1p2 che del genoma e come sonda di ibridazione il clone in esame. Dalla FIGURA 38 si evidenzia che il dot della sonda con concentrazione pari a 0,125 ng/μl ha una buona corrispondenza con il dot del genoma con concentrazione pari a 100 ng/μl e pertanto la sequenza è presente nel genoma con una percentuale pari allo 0,125%.

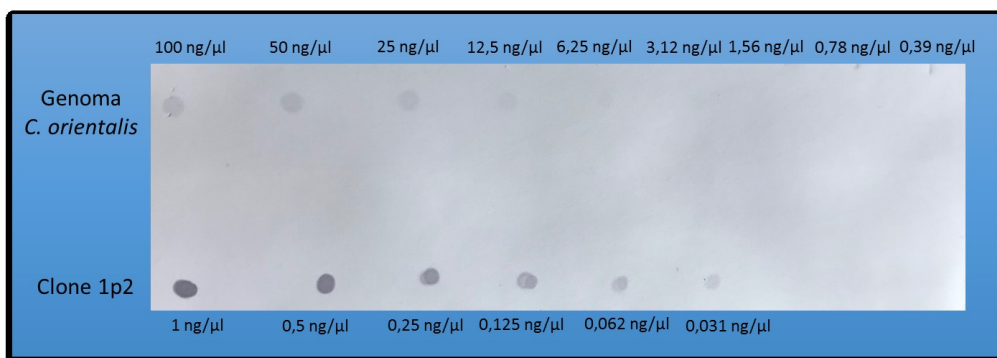


FIGURA 38 Dot blot quantitativo della sequenza del clone 1p2.

3.1.2. Risultati della sequenza inserto del clone 7p1

La sequenza dell'inserto presenta una lunghezza di 847 bp e ha un contenuto in percentuale di nucleotidi GC del 49,47%.

È stata condotta l'analisi della sequenza su Censor dove sono stati rilevati due regioni LTR-DIRS: DIRS-6_XT, una sequenza lunga 76-390 bp, contenuta nella regione iniziale della sequenza; DIRS-17_Psi, una sequenza lunga circa 686-813 bp contenuta nella porzione terminale (FIGURA 109).

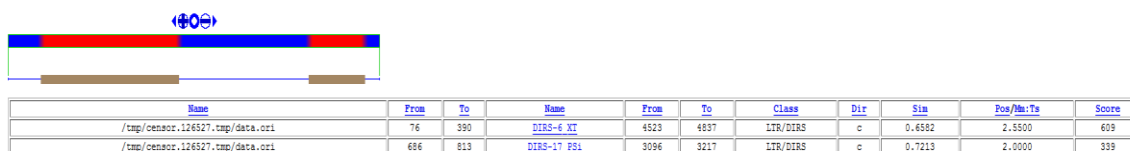


FIGURA 19 Analisi su Censor della sequenza del clone 7p1.

Dato che la sequenza di DIRS-6_XT è più lunga ed appartiene a *X. tropicalis*, ne è stata ricavata la sequenza nucleotidica che a sua volta è stata analizzata su BlastX. I risultati hanno mostrato un'identità del 54,81% con la sequenza di una proteina non identificata in *X. tropicalis* (KAE8606559.1). Quest'ultima è stata cercata anche su BlastP e si è visto che la sequenza proteica presenta un dominio C-terminale per enzimi con sito catalitico attivo della superfamiglia di DNA-breaking rejoining che includono topoisomerasi di tipo IB e anche tirosina ricombinasi, enzima caratteristico della regione C-terminale dei DIRS (FIGURA).

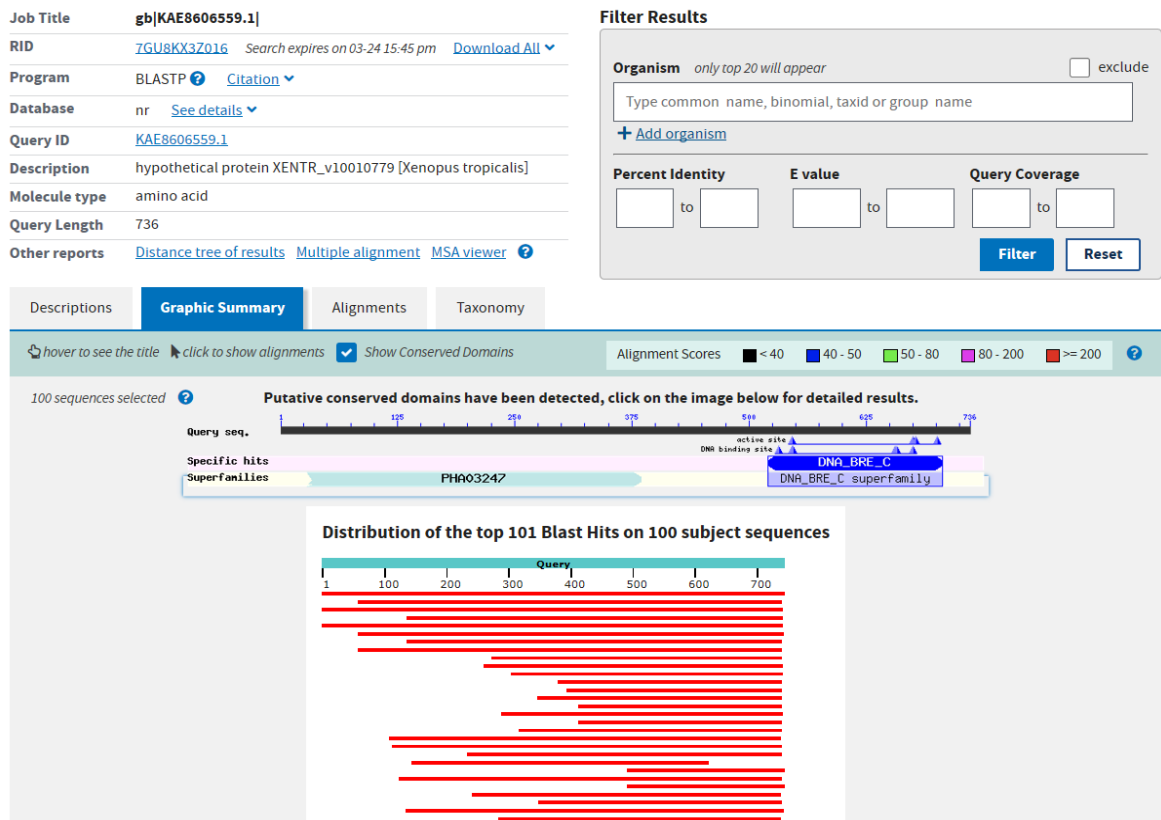


FIGURA 20 Analisi su BlastP della sequenza di *X. tropicalis* ottenuta da Censor usando il clone cl7p1.

L'analisi della sequenza inserito del cl7p1 è stata effettuata anche su ORFfinder, dove sono stati individuati tre ORF: ORF13, ORF15 e ORF17, sovrapposti alla regione che maschera il DIRS trovato su Censor (FIGURA 21).

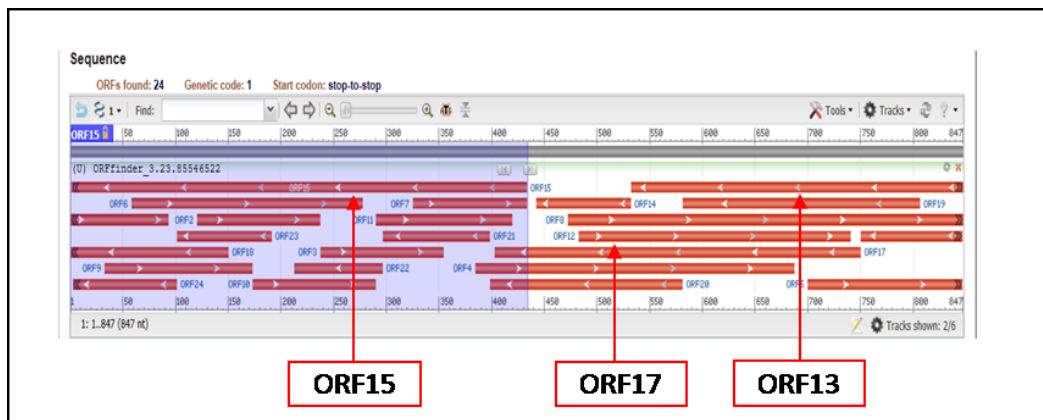


FIGURA 21 Risultato dell'analisi condotta con ORF finder sulla sequenza inserito del clone 7p1. Le sequenze indicate sono le ORF13, ORF15 e ORF17.

La sequenza dell'ORF15 è stata analizzata su BlastP ed ha una buona identità con proteine non caratterizzate di diversi organismi, tra cui lo *X. tropicalis*, ed in particolare con la sequenza individuata nell'analisi precedente (KAE8606559.1) sulla sequenza DIRS-6_XT e anche con varie isoforme di RNA-elicasi in *X. laevis*.

Anche la sequenza dell'ORF13 è stata analizzata su BlastP ed ha una buona identità con sequenze di proteine non identificate, ma il dominio conservato è quello della regione della retrotrascrittasi (FIGURA).

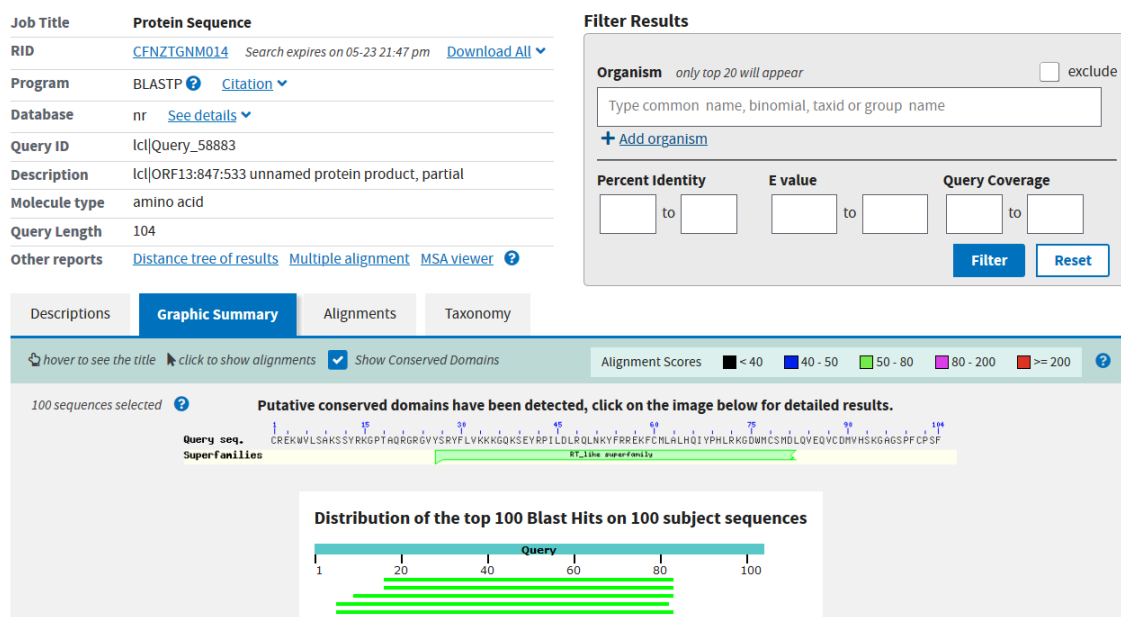


FIGURA 22 Analisi su BlastP della sequenza dell'ORF13.

Considerando il dominio conservato sul database CDD e poi anche su CPDF è stato visto che in molte piante e in molti animali questo dominio resta attivo (FIGURA).



FIGURA 23 Analisi dei siti attivi della regione della retrotrascrittasi analizzata dalla sequenza di ORF13.

Per ultima anche la sequenza dell'ORF17 è stata analizzata su BlastP ed ha una bassa identità con proteine non caratterizzate di pesci e anfibi, ma ha anche un'identità del 44% con il dominio della tirosina ricombinasi/integrasi di un batterio, *Nocardioides jejuensis* (FIGURA).

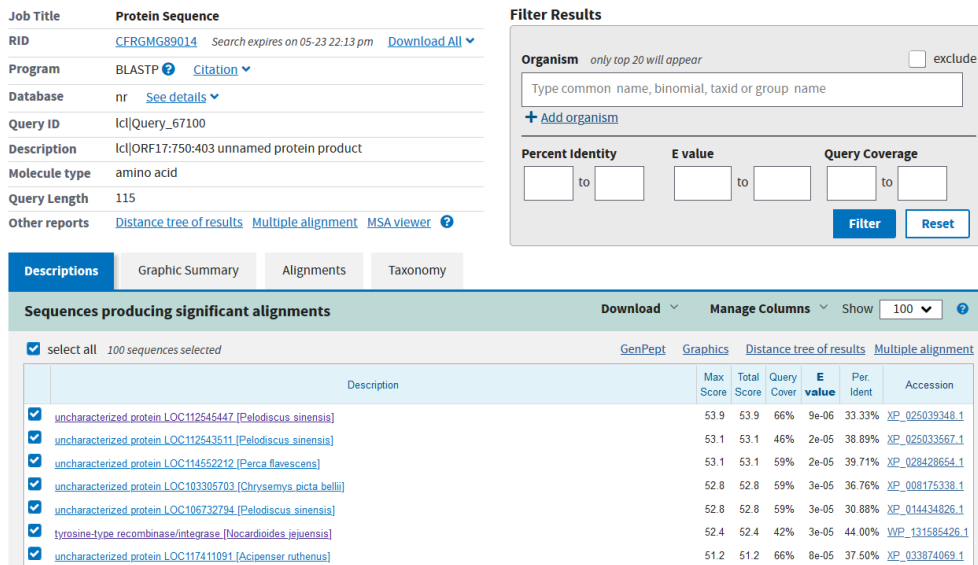


FIGURA 24 Analisi su BlastP della sequenza dell'ORF13.

Attraverso la caratterizzazione di queste tre sequenze ORF, individuate nella sequenza inserto del clone 7p1, è stato possibile ricostruire la composizione dei domini caratteristici degli elementi DIRS.

Si è proseguito con l'analisi sul trascrittoma di *C. orientalis* a partire dalla sequenza inserto del clone 7p1. Il risultato ha permesso di identificare diverse sequenze del trascrittoma, tra cui quelle considerate nell'allineamento qui di seguito.

Allineamento del clone 7p1 con i primi tre risultati ottenuti dall'analisi condotta sul trascrittoma

TRINITY_DN92999_c4_g1_i3	-----	0
Clone7p1	-----GGTAGAAGGCAAGCGCACGTTTT	23
TRINITY_DN108654_c7_g1_i12	AGTTGGTCATCCACCCGGAAGTGTGTGCGCTGCAGATAGAAAGACAGCGCACGTTTT	60
TRINITY_DN92999_c3_g1_i4	-----	0
TRINITY_DN92999_c4_g1_i3	-----	0
Clone7p1	GGGTCCTGACGGTGGAGCCTCTCTCCTCTTTTGGAGGAAGCGGAGGTGAAAGAAGGTG	83
TRINITY_DN108654_c7_g1_i12	GGGTCCAGACGGTGGAGCCTCTCTCCTCTCTAGAAAGGATGCGGAGGCGGAAAGAAGGTC	120
TRINITY_DN92999_c3_g1_i4	-----	0
TRINITY_DN92999_c4_g1_i3	-----	0
Clone7p1	GGAAAAGTGATGGACTGCTCAAGATGGAAGGGCGTGACATCCTTTGGCAAGAAGGAAGAA	143
TRINITY_DN108654_c7_g1_i12	GGCAAGGTGATGGTCTGCCCCAGGTGGAAGGGGTGACAACTTTGGGCAAGAAAGAAGGA	180
TRINITY_DN92999_c3_g1_i4	-----	0
TRINITY_DN92999_c4_g1_i3	-----	0
Clone7p1	CGAGTCTCAAGACCACCTTGTCCGGATGAACAGAAAGGAAGGAGGTTCAGAAGAGAGG	203
TRINITY_DN108654_c7_g1_i12	CGAGTCTTTAGGACTACTTTGTCCGGATGGACAGAGAGAAAGGAGGTTTACAGGAAAGA	240
TRINITY_DN92999_c3_g1_i4	-----	0
TRINITY_DN92999_c4_g1_i3	--TTGCAGCTCACTCACCTGCGCGCAGAAGTGATGGCCACCAGAAAGGCAGTTTGTAGG	58
Clone7p1	GCCTGGAGCTCACTCACTCTTCTGGCAGAGGTGATGGCCACCAGGAAAGCAGTCTTTCAGA	263
TRINITY_DN108654_c7_g1_i12	GCCTGGAGCTCACTCACTCTTCTGGCAGAGGTGAGGGCTACCAGGAAAGCAGTCTTTAGA	300
TRINITY_DN92999_c3_g1_i4	-----CTTTAGA	7
	** **.	
TRINITY_DN92999_c4_g1_i3	GTTAAAAGTCTCAGAGGACAGTTATGTAGCGTCTCAAAGGGAGTGACACATAAGATAGGTA	118
Clone7p1	GTCAGAAGACGTAGAGGACACTTGTGAAGTGGCTCAAAGAGAGCACACATCAGGAATGTT	323
TRINITY_DN108654_c7_g1_i12	GTAAGAAGACGAAGAGGACAGTTATGAAGGGGCTCAAAGGGAGCACACATCAGATAAGTC	360
TRINITY_DN92999_c3_g1_i4	GTCAGAAGACGTAGAGGACAGTTGTGAAGTGGCTCAAAGGGAGCACACATCAGATATGTT	67
	** * .***.* ***** **.*:** * *****.*** .*****.***.* **	
TRINITY_DN92999_c4_g1_i3	AGAACTAAGTTCAAGTCCCCTGAGGCATTACGAAGGGGAACAGGAGGAAAGAGATGAGTA	178
Clone7p1	AGAACAAGGTTCAAGTCCCCTGAGGCATTACGAAGGGGAACAGGAGGAAACAAATGTGTG	383
TRINITY_DN108654_c7_g1_i12	AAGACAAGGTTCAAGTCCCCTGAGGCATTACGAAGGGTACAGGTGGAACATGTGAGTG	420
TRINITY_DN92999_c3_g1_i4	AAGACAAGGTTCAAGTCCCCTGAGGCATTACGAAGGGTACAGGTGGAACATGTGAGTG	127
	*.***.*.*****.***** **.*:**.*****.***** * .**:**.	

TRINITY_DN92999_c4_g1_i3	ATGCCCTTAAGAAAGCGATTACAAACAGGGGACTTGAAGAGAGAAGGCTGATCAGGAAGC	238
Clone7p1	ATGCCCTTCAAGAATCTATTTACA-ACGGGGACTTGAAGAGAGAAGGCTGATCAGGAAGC	442
TRINITY_DN108654_c7_g1_i12	ATGCCCTTAAAGAACCTATTTACAATCGGGGACTTGAATAAGGAAGGTTGATCAGGTAGC	480
TRINITY_DN92999_c3_g1_i4	ATGCCCTTAAAGAACCTATTTACAATCGGGGACTTGAATAAGGAAGGTTGATCAGGTAGC	187
	***** *..* * ***** .***** *..***** *****:***	
TRINITY_DN92999_c4_g1_i3	CTAAGAAAGGCTGAAATGGCAGAGAGGTAGCCCTTTAGAGTGCCAGAGCAGAGCCCTGC	298
Clone7p1	CTAAGAAAGGCAGAAATAGCCGAGAGATAGCCTTTGATAGTGCCAGAGCAGAGCCCTGC	502
TRINITY_DN108654_c7_g1_i12	CTTAGAAAGGCTGAAATGGCTGATAGATAACCTTTAAGAGTGCCAGAGCAGAGCCCTGC	540
TRINITY_DN92999_c3_g1_i4	CTTAGAAAGGCTGAAATGGCTGAGAGGTAACCTTTGATAGTGCCAGAGCAGAGCCCTGC	247
	.:***:*****.* * ** *.**.* ** * * *****:*****	
TRINITY_DN92999_c4_g1_i3	TGGGCCAAAGAAAGGAGGAACAAGAGAACCTCAGAAAGAGGGGCAGAAAGGGGATCCACA	358
Clone7p1	TGGGCCAAAGAAAGAAGGAACAAGAGACCTCAGAAAGAGGGGCAGAAAGGGGATCCAGC	562
TRINITY_DN108654_c7_g1_i12	TGGGCCAAAGAGAGAAGAAAAGAAGAACTCTGAAAGAGGGGCAGAAAGGGGATCCACA	600
TRINITY_DN92999_c3_g1_i4	CGGGCTAAAGAGACAAGAAAAGA-----	271
	**** *.*.*.*.*.*..	
TRINITY_DN92999_c4_g1_i3	GCCTTTGCTGTGCA-----	372
Clone7p1	TCCTTTGCTGTGCACCATGTCACAACTTGTTCACCTGCAGATCCATCGAACACATCCA	622
TRINITY_DN108654_c7_g1_i12	GCCTTTGCTGTGCACCATGCCACAAATTTCTTCCACCTGCAGGCGTATA-----CCG	652
TRINITY_DN92999_c3_g1_i4	-----	271
TRINITY_DN92999_c4_g1_i3	-----	372
Clone7p1	ATCCCTTTGCGAAGGTGCGGATATATTTGGTGGAGAGCCAGCATGCAAAAT----TTCT	678
TRINITY_DN108654_c7_g1_i12	A----TTTAG-----TGGAGGACGCCTGGCTGCCAGAATGACATCA	690
TRINITY_DN92999_c3_g1_i4	-----	271
TRINITY_DN92999_c4_g1_i3	-----	372
Clone7p1	C--TCGCTGAAGT---ACTT---GTTGAGCTGTGCAAGATC-----	712
TRINITY_DN108654_c7_g1_i12	CAGACTTCGGGAGGCAAGTCAAATGGAATCAACTGTCGCCGCTCAATCTCCACGCATGGA	750
TRINITY_DN92999_c3_g1_i4	-----	271
TRINITY_DN92999_c4_g1_i3	-----	372
Clone7p1	-----CAAATGGGTCTGTACTCCGATTTCTGACCTTTTTTCTTT	752
TRINITY_DN108654_c7_g1_i12	GGTGGAGGCTGGACAGGTTCCGGTGGAGGACCGTCCCTG-----CTGCTGC	797
TRINITY_DN92999_c3_g1_i4	-----	271
TRINITY_DN92999_c4_g1_i3	-----	372
Clone7p1	ACCAGAAAATATCTGGAATACACTCCTCGACCTCTTTAGCCGTGGGACCTTTTCTATAG	812
TRINITY_DN108654_c7_g1_i12	GACAGAAGAT-----CCTCCCGAAGGGGCAGACGAATCGGGGGATCTGTACCATG	848
TRINITY_DN92999_c3_g1_i4	-----	271
TRINITY_DN92999_c4_g1_i3	-----	372
Clone7p1	CTGCTTTTCGCAGA-----	826
TRINITY_DN108654_c7_g1_i12	CTGAGTAGCTCGGGATACCAGAATCTTCGTGCCAGTCCGGAGCCACAAGAATGACTTGG	908
TRINITY_DN92999_c3_g1_i4	-----	271
TRINITY_DN92999_c4_g1_i3	-----	372
Clone7p1	-----GAGCAC-----CCTCTCTCTG--CA-----	847
TRINITY_DN108654_c7_g1_i12	GCCCGGTCGCGCCGAATCTTCTCAGCACTCGGGGCAGAAGTGGTATAGCGGAAAGGCC	968
TRINITY_DN92999_c3_g1_i4	-----	271
TRINITY_DN92999_c4_g1_i3	-----	372
Clone7p1	-----	847
TRINITY_DN108654_c7_g1_i12	TACAACAGACCGCA	982
TRINITY_DN92999_c3_g1_i4	-----	271

In particolare la sequenza TRINITY_DN108654_c7_g1_i12 è stata utilizzata per effettuare ulteriori analisi dato che presenta una lunghezza di 982 bp ed è

quindi maggiore rispetto alla sequenza del clone 7p1, con cui condivide una regione conservata, come è possibile verificare nell'allineamento di seguito.

```

Clone7p1          -----GGTAGAAGGCAAG
TRINITY_DN108654_c7_g1_i12  AGTTGGTCATCCACCCGGAAGTATGTGTGCGCTGCAGATAGAAAGACAG
                                *  *  *  *  *  *  *  *

Clone7p1          CGCACGTTTTGGGTCTGACGGTGGAGCCTCTCTTCCTCTTTTGAGGAAG
TRINITY_DN108654_c7_g1_i12  CGCACGTTTTGGGTCCAGACGGTGGAGCCTCTCTTCCTCTCTAGAAGGAT
                                *****
                                ***** *  *  *  *

Clone7p1          GCGGAGGTGAAAGAAGGTGGGAAAAGTGTGACTGCTCAAGATGGAAG
TRINITY_DN108654_c7_g1_i12  GCGGAGGCGAAAGAAGGTGCGCAAGGTGATGGTCTGCCCCAGGTGGAAG
                                ***** *  *  *  *  *  *  *  *

Clone7p1          GGCCTGACATCCTTTGGCAAGAAGGAAGAAGAGTCTCAAGACCACCTT
TRINITY_DN108654_c7_g1_i12  GGGGTGACAACTTTGGGCAAGAAAGAAGGACGAGTCTTTAGACTACTTT
                                ** ***** *  *  *  *  *  *  *

Clone7p1          GTCCGGATGAACAGAAAGGAAAGGAGGTTTACAGGAAAGAGCCTGGAGCT
TRINITY_DN108654_c7_g1_i12  GTCCGGATGGACAGAGAGGAAAGGAGGTTTACAGGAAAGAGCCTGGAGCT
                                ***** *  *  *  *  *  *  *

Clone7p1          CACTCACTCTTCTGGCAGAGGTGATGGCCACCAGGAAAGCAGTCTTCAGA
TRINITY_DN108654_c7_g1_i12  CACTCACTCTTCTGGCAGAGGTGAGGGCTACCAGGAAAGCAGTCTTTAGA
                                ***** *  *  *  *  *  *  *

Clone7p1          GTCAGAAGACGTAGAGGACACTTGTGAAGTGGCTCAAAGAGAGCACACAT
TRINITY_DN108654_c7_g1_i12  GTAAGAAGACGAAGAGGACAGTTATGAAGGGGCTCAAAGAGGAGCACACAT
                                ** ***** *  *  *  *  *  *  *

Clone7p1          CAGGAATGTTAGAACAAGGTTCAAGTCCCCTGAGGCATGACGAAAGGAA
TRINITY_DN108654_c7_g1_i12  CAGATAAGTCAAGACAAGGTTCAAGTCCCCTGAGGCATGACGAAAGGAA
                                *** *  *  *  * ***** *  *  *  *

Clone7p1          CAGGAGGAAACAATGTGTGATGCCCTTCAAGAATCTATTTACAA-CGGG
TRINITY_DN108654_c7_g1_i12  CAGGTGGAACATGTGAGTGTGATGCCCTTTAAGAACCTATTTACAATCGGG
                                **** ***** *  * ***** *  * ***** *  *

Clone7p1          GACTTGAAGAGAGAAGGCTGATCAGGAAGCCTAAGAAAGGCAGAAATAGC
TRINITY_DN108654_c7_g1_i12  GACTTGAATAAGGAAGGTTGATCAGGTAGCCTTAGAAAGGCTGAAATGTC
                                ***** *  *  *  *  *  *  *  *  *  *  *

Clone7p1          CGAGAGATAGCCTTTGATAGTGCCAGAGCAGAGCCCTGCTGGGCCAAAG
TRINITY_DN108654_c7_g1_i12  TGATAGATAACCTTTAAGAGTGCCAGAGCAGAGCCCTGCTGGGCCAAAG
                                ** ***** *  * *****

Clone7p1          AAAGAAGGAACAAAAGGACCTCAGAAAGAGGGGCAGAAAGGGGATCCAGC
TRINITY_DN108654_c7_g1_i12  AGAGAAGAAAAGAAGAACCCTCTGAAAGAGGGGCAGAAAGGGGATCCACA
                                *  *  *  *  *  * *****

Clone7p1          TCCTTTGCTGTGCACCATGTCACAACTTGTTCACCTGCAG-----AT
TRINITY_DN108654_c7_g1_i12  GCCTTTGCTGTGCACCATGCCACAAATTTCTTCACCTGCAGGGCGTATAC
                                ***** *  * ***** *

Clone7p1          CCATCGAACACATCCAATCCCCTTTGCGAAGGTG----CGGATATATTTG
TRINITY_DN108654_c7_g1_i12  CGATTTAGTGGAGGGACGCTGGCTGCCAGAATGACATCACAGACTTCGG
                                *  *  *  *  *  *  *  *  *  *  *  *  *

Clone7p1          GTGGAGAGCCA---GCATGCAA-----AATTTCTCTCGTCTGA

```



```

TRINITY_DN108654_c7_g1_i12      GAGGCAAGTCAAATGGAATCAACTGTCGCCGCTCAATCTCCACGCATGGA
* * * * * * * * * * * * * * * * * * * * * * * * * * *
Clone7p1
TRINITY_DN108654_c7_g1_i12      AGTACTTGTGTGAGCTGTCTCGAAGATCCAAAATGGGTCTGTACTCCGATTTCT
GGTGGAGGCTGGACAGGTTCTGGGTGGAGGACCGTCCCGTGTCTGCGCAC
* * * * * * * * * * * * * * * * * * * * * * * * * * *
Clone7p1
TRINITY_DN108654_c7_g1_i12      TGACCTTTTTTCTTTACCAGAAAAT--ATCTGGAATACACTCTCTCGACCT
AGAAGATCTCCCGAAGGGGCAGACGAATCGGGGATCTGTACCATGCT
* * * * * * * * * * * * * * * * * * * * * * * * * * *
Clone7p1
TRINITY_DN108654_c7_g1_i12      CTTTGAGCCGTGGGACCT---TTTCTATAGCTGCTTT----TCGCAGAGA
GAGTAGCTCGGGATACCAGAAATCTTCGTGCCAGTCCGGAGCCACAAGAA
* * * * * * * * * * * * * * * * * * * * * * * * * * *
Clone7p1
TRINITY_DN108654_c7_g1_i12      GCACCCACTTCTCTCTGCA-----
TGACTTGGGCCCGGTCGCGCCGAATCTTCTTCAGCACTCGGGGCAGAAGT
* * * * * * * *
Clone7p1
TRINITY_DN108654_c7_g1_i12      -----
GGTATAGCGGAAAGGCGTACAACAGACCGCA

```

Dapprima la sequenza TRINITY_DN108654_c7_g1_i12 è stata analizzata su Censor da cui risulta la corrispondenza completa con DIRS-2a_Lch (FIGURA).



FIGURA 25 Analisi su Censor della sequenza TRINITY_DN108654_c7_g1_i12 ottenuta dal trascrittoma di *C. orientalis*.

La sequenza è stata analizzata anche su ORFfinder dove l'ORF22 si sovrappone completamente (FIGURA).

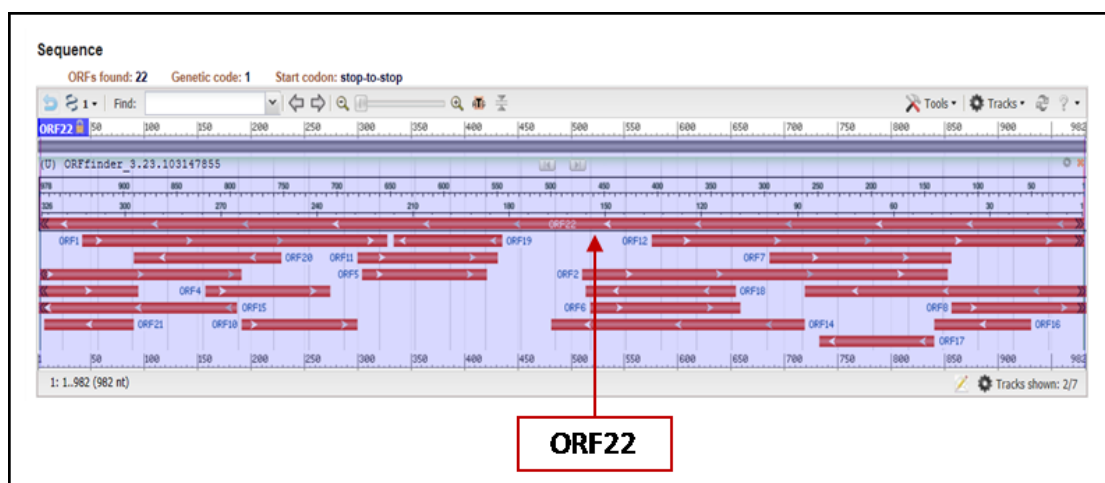


FIGURA 26 Analisi su ORFfinder della sequenza TRINITY_DN108654_c7_g1_i12.

Per cui si è proceduto all'analisi dell'ORF22 su BlastP ed il risultato ottenuto è lo stesso di quello ottenuto con l'ORF15 della sequenza inserto. Per un'ulteriore verifica è stato effettuato l'allineamento tra le due ORF, a dimostrazione del fatto che si tratta della stessa regione.

Allineamento della sequenza del ORF15 con ORF22.

```

lc1|ORF22      RSVVRLSAYTTSAPSAEEDSARPGPSHSCGSLGTLKILVSRATQHGDSPDSSAPSGGSS      60
lc1|ORF15      -----
lc1|ORF22      VAAAGDGPPEPVQPPPPCVEIERRQLIPFDLPPEVCDVILAARRPSTKSVYACRWKKFV      120
lc1|ORF15      -----
lc1|ORF22      AWCTAKAVDPLSAPLSEVLLFLLSLAQGSALGTLKGYLSAISAFRLRPDQPSLFKSPFV      180
lc1|ORF15      -----MAFSLQVPEVV
                                                :: **
lc1|ORF22      NRFLKGITHMFPVPFVMPQWDLNLVLTLYLMCAPFEPLHNCPLRLLTLKTAFLVAITSAR      240
lc1|ORF15      NRFLKGITHLFPPVPFVMPQWDLNLVLTFLMCALFEPLHKCPLRLLTLKTAFLVAITSAR      70
*****.:*****.:**** *****.:*****.:****
lc1|ORF22      RVSELQALSCKPPFLSVHPDKVVLKTRPSFLPKVVTFFHLGQTITLPTFFPPPHPSREEE      300
lc1|ORF15      RVSELQALSSEPPFLSVHPDKVVLKTRSSFLPKDVTFFHLEQSITFFPTFFPPPPSSKEEE      130
*****.:*****.:** ***** ***** *.:**.****** *.:***
lc1|ORF22      RLHRLDPKRALSFYLQORTHQFRVDDQ      326
lc1|ORF15      RLHRQDPKRALAFY-----      144
**** *****.:**

```

È stato infine valutato quanto l'elemento analizzato sia presente nel genoma di *C. orientalis* attraverso un dot blot quantitativo, utilizzando concentrazioni scalari sia della sequenza cl7p1 che del genoma e come sonda di ibridazione il clone in esame. Dalla *FIGURA* si evidenzia che il dot della sonda con concentrazione pari a 0,031 ng/μl ha una buona corrispondenza con il dot del genoma con concentrazione pari a 100 ng/μl e pertanto la sequenza è presente nel genoma con una percentuale pari allo 0,031%.

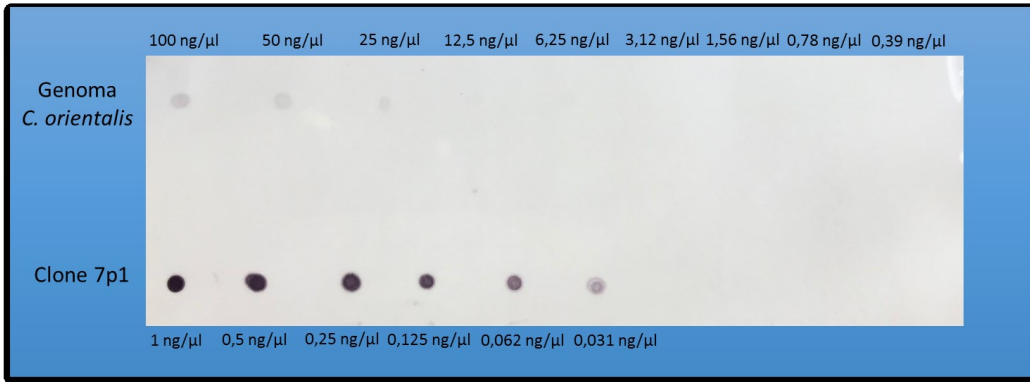


FIGURA 27 Dot blot quantitativo della sequenza del clone 7p1.

3.1.3. Risultati della sequenza inserto del clone 59p1

La sequenza inserto identificata nel clone 59p1 è lunga 668 bp ed ha una composizione in GC del 48,80%.

L'analisi su Censor ha messo in evidenza la presenza di una sequenza ripetuta riconducibile ad un DNA satellite, ATSAT2, lunga 199 bp nella parte terminale dell'inserto (FIGURA).



FIGURA 28 Risultato dell'analisi della sequenza inserto cl59p1 su Censor.

Avendo individuato una sequenza possibilmente ripetuta in tandem, è stata effettuata l'analisi su DotLet che ha evidenziato la presenza di due ripetizioni, rappresentate nel grafico dalle due linee diagonali più corte (FIGURA).

Allineamento dei due tratti ripetuti.

```

Rip1      AAGCAAAACCAAAGGGGATGATGGACGGAGTGTTGAACATTGGAAGCACTCACCCCTAGT
Rip2      AAGCAAAACCAAAGGGGATGATGGAGGGAGTGTTGAACATTGCAAGCACTCACCCCCAGT
          *****

Rip1      CAACTGATCTGGGTTTAATCCATCATCATTCTTTAGCTCACTCTACCGCCCCAGTTCAGA
Rip2      CAACTGATCTGGGTTTAATCCATCATCATTCTTTATCTCACTCTACCAACCCAGTTCAGA
          *****

```

Rip1
Rip2

AATGGCCATATGCAAATCAGTCTGACCTGTGCCTCATG
CATGGCCATATGCAAATCAGTCTGACCTGTGCCTCATG

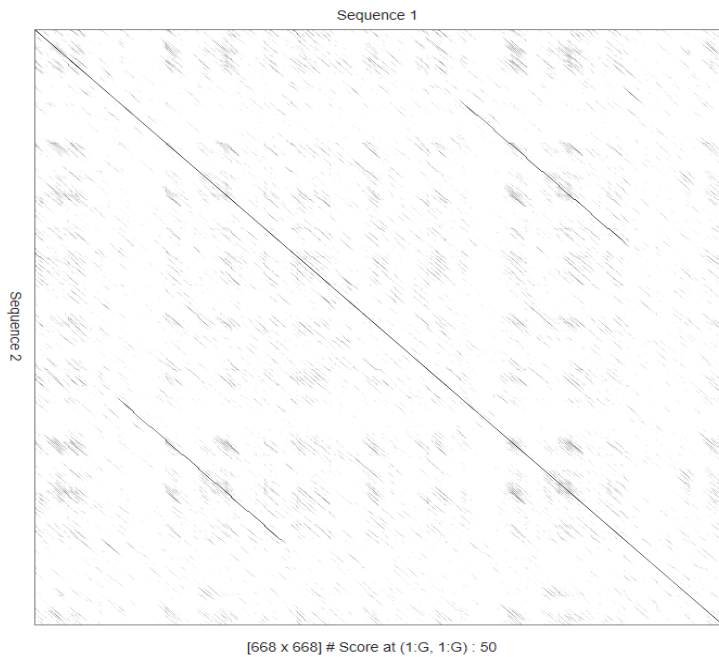


FIGURA 29 Analisi della sequenza del clone 59p1 su DotLet.

È stata eseguita l'analisi su BlastN (FIGURA 30) che ha messo in evidenza l'identità del 96% con una sequenza satellite di tipo 2 individuata in *Cynops pyrrhogaster*. La sequenza è lunga 105 bp e si allinea con la sequenza inserto nella regione centrale.

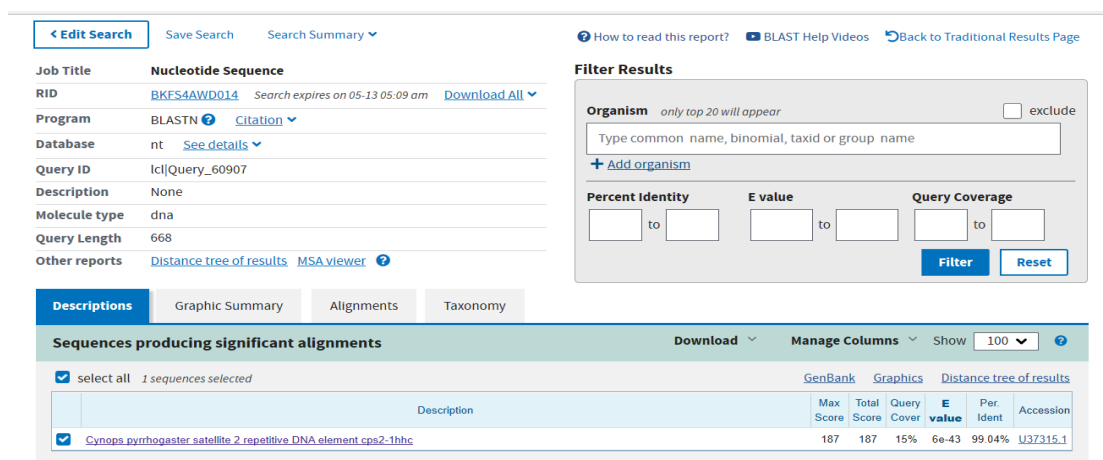


FIGURA 30 Analisi su BlastN della sequenza del clone 59p1.

Continuando l'analisi su BlastN abbassando la stringenza di identità sono state individuate sequenze che presentano una buona percentuale di identità con la sequenza inserto c159p1. Infatti oltre alla sequenza di *C. pyrrhogaster* ci sono le sequenze di altre salamandre, come quelle di *Ambystoma talpoideum*, ma anche di tritoni come *Notophthalmus viridescens* e *Lissotriton vulgaris* (FIGURA 5).

Job Title	Nucleotide Sequence
RID	CWRT55XJ016 Search expires on 05-28 20:37 pm Download All ▼
Program	BLASTN Citation ▼
Database	nt See details ▼
Query ID	lcl Query_23255
Description	None
Molecule type	dna
Query Length	668
Other reports	Distance tree of results MSA viewer ?

Filter Results

Organism only top 20 will appear exclude

[+ Add organism](#)

Percent Identity to **E value** to **Query Coverage** to

[Filter](#) [Reset](#)

Descriptions
Graphic Summary
Alignments
Taxonomy

Sequences producing significant alignments Download ▼ Manage Columns ▼ Show 100 ▼ [?](#)

select all 37 sequences selected [GenBank](#) [Graphics](#) [Distance tree of results](#)

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/> Notophthalmus viridescens satellite 2 DNA	253	423	63%	7e-63	83.40%	X04478.1
<input checked="" type="checkbox"/> Cynops pyrrhogaster satellite 2 repetitive DNA element cps2-1hhc	184	184	15%	1e-41	99.04%	U37315.1
<input checked="" type="checkbox"/> Cynops pyrrhogaster ghr mRNA for growth hormone secretagogue receptor 1a, complete cds	126	182	42%	3e-24	74.87%	LC050183.1
<input checked="" type="checkbox"/> Cynops pyrrhogaster satellite 2 repetitive DNA element cps2-2hhc	117	117	15%	1e-21	85.29%	U37317.1
<input checked="" type="checkbox"/> Salamandrina perspicillata microsatellite SAL06 sequence	85.1	136	28%	8e-12	76.72%	KT373916.1
<input checked="" type="checkbox"/> Lissotriton vulgaris ampelensis clone INCLva-5 microsatellite sequence	84.2	144	48%	8e-12	77.39%	JQ320452.1
<input checked="" type="checkbox"/> Ambystoma talpoideum satellite 2 repetitive element, clone pAtal 1	68.9	68.9	10%	6e-07	84.06%	L05855.1
<input checked="" type="checkbox"/> Ambystoma talpoideum satellite 2 repetitive element, clone pAtal 5	66.2	66.2	10%	2e-06	82.86%	L05867.1
<input checked="" type="checkbox"/> Ambystoma talpoideum satellite 2 repetitive element, clone pAtal 4	66.2	66.2	10%	2e-06	82.86%	L05864.1
<input checked="" type="checkbox"/> Ambystoma talpoideum satellite 2 repetitive element, clone pAtal 2	66.2	66.2	10%	2e-06	82.86%	L05861.1
<input checked="" type="checkbox"/> Ambystoma talpoideum satellite 2 repetitive element, clone Atalcon	66.2	66.2	10%	2e-06	82.86%	L05859.1
<input checked="" type="checkbox"/> Ambystoma talpoideum satellite 2 repetitive element, clone pAtal 3	66.2	66.2	10%	2e-06	82.86%	L05857.1
<input checked="" type="checkbox"/> Amphiuma tridactylum satellite 2 repetitive element, clone pAtri D	59.0	59.0	13%	3e-04	75.82%	L05841.1
<input checked="" type="checkbox"/> Ambystoma talpoideum satellite 2 repetitive element, clone pAtal 12	58.1	58.1	9%	0.001	80.30%	L05858.1

FIGURA 5 Analisi della sequenza del clone 59p1 su BlastN utilizzando una minore stringenza di identità.

È evidente che questa sequenza, anche in tratti brevi, sia diffusa tra i Caudata, come si vede nell'allineamento della sequenza inserito con i risultati ottenuti su BlastN (FIGURA).

Distribution of the top 42 Blast Hits on 37 subject sequences

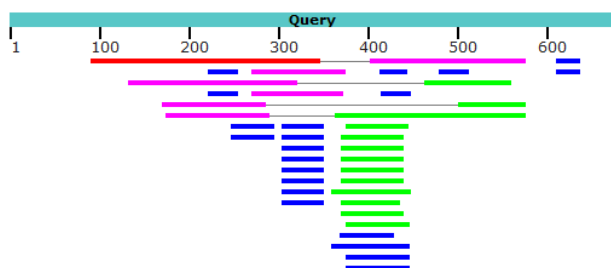


FIGURA 32 Risultato dell'analisi BlastN della sequenza del clone 59p1 con sequenze di altre salamandre e tritoni.

La sequenza inserito è stata cercata anche nei genomi sequenziati di anuri (FIGURA), di caudati (FIGURA) e dei gimnofioni. Tra i risultati più rilevanti si ha un'identità della sequenza con quella di *Nanorana parkeri* e con diverse sequenze ritrovate in *Ambystoma mexicanum*, mentre non sono stati rilevati risultati nei cecilidi.

Job Title Nucleotide Sequence

RID BWTd4K27016 Search expires on 05-16 17:49 pm [Download All](#)

Program BLASTN [Citation](#)

Database wgs (15 databases) [See details](#)

Query ID lcl|Query_16699

Description None

Molecule type dna

Query Length 668

Other reports [Distance tree of results](#) [MSA viewer](#)

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity to

E value to

Query Coverage to

[Filter](#) [Reset](#)

Descriptions [Graphic Summary](#) [Alignments](#)

Sequences producing significant alignments [Download](#) [Manage Columns](#) Show

select all 100 sequences selected

	Description	Max Score	Total Score	Query Cover	E value	Per Ident	Accession
<input checked="" type="checkbox"/>	Nanorana parkeri isolate BGI_ZX_2015 contig142157_w... whole genome shotgun sequence	226	356	55%	1e-56	83.26%	JYQUB1142156.1
<input checked="" type="checkbox"/>	7612269_w... whole genome shotgun sequence	73.4	73.4	7%	2e-10	92.00%	VKOB011013150.1
<input checked="" type="checkbox"/>	7076843_w... whole genome shotgun sequence	71.6	71.6	8%	5e-10	88.33%	VKOB010745437.1
<input checked="" type="checkbox"/>	5594574_w... whole genome shotgun sequence	71.6	71.6	8%	5e-10	88.33%	VKOB010022799.1
<input checked="" type="checkbox"/>	5274544_w... whole genome shotgun sequence	71.6	71.6	8%	5e-10	88.33%	VKOB010013600.1

FIGURA 33 Analisi della sequenza cl59p1 su BlastN, utilizzando il database whole-genome shotgun contigs (WGS) e restringendo la ricerca agli Anura.

Job Title Nucleotide Sequence

RID [C4F1HN89014](#) Search expires on 05-19 15:41 pm [Download All](#) ▼

Program BLASTN [Citation](#) ▼

Database wgs (98 databases) [See details](#) ▼

Query ID Icl|Query_42835

Description None

Molecule type dna

Query Length 668

Other reports [Distance tree of results](#) [MSA viewer](#) ?

Filter Results

Organism *only top 20 will appear* exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity to **E value** to **Query Coverage** to

[Filter](#) [Reset](#)

Descriptions | Graphic Summary | Alignments

Sequences producing significant alignments [Download](#) ▼ [Manage Columns](#) ▼ Show 100 ▼ ?

select all 100 sequences selected [GenBank](#) [Graphics](#) [Distance tree of results](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Ambystoma mexicanum strain DD151 AMEXG_0030009480_whole genome shotgun sequence	100	600	27%	1e-17	73.82%	PGSH01098165.1
<input checked="" type="checkbox"/>	Ambystoma mexicanum strain DD151 AMEXG_0030008879_whole genome shotgun sequence	100	3065	35%	1e-17	71.06%	PGSH01096902.1
<input checked="" type="checkbox"/>	Ambystoma mexicanum strain DD151 AMEXG_0030003170_whole genome shotgun sequence	99.6	800	39%	4e-17	73.30%	PGSH01085532.1
<input checked="" type="checkbox"/>	Ambystoma mexicanum strain DD151 AMEXG_0030004957_whole genome shotgun sequence	99.6	258	36%	4e-17	72.87%	PGSH01069209.1
<input checked="" type="checkbox"/>	Ambystoma mexicanum strain DD151 AMEXG_0030122817_whole genome shotgun sequence	99.6	358	36%	4e-17	72.00%	PGSH01020099.1
<input checked="" type="checkbox"/>	Ambystoma mexicanum strain DD151 AMEXG_0030000155_whole genome shotgun sequence	98.7	747	34%	4e-17	71.95%	PGSH01067904.1

FIGURA 34 Analisi della sequenza *cl59p1* su BlastN, utilizzando il database (WGS) e restringendo la ricerca ai *Caudata*.

In parallelo è stata eseguita un'analisi sul trascrittoma di *C. orientalis*, dove sono state individuate sei sequenze. Considerando le sequenze del trascrittoma e quelle ottenute con la ricerca precedente è stata effettuata un'analisi filogenetica (FIGURA 6).

Tale analisi evidenzia due raggruppamenti: il primo contiene sequenze appartenenti principalmente ad *Ambystoma* di varie specie; il secondo contiene sequenze appartenenti a varie specie di *Cynops*, tra cui quelle individuate nel trascrittoma, ma anche *Notophthalmus viridescens*, *Triturus vulgaris* e *Nanorana parkeri*.

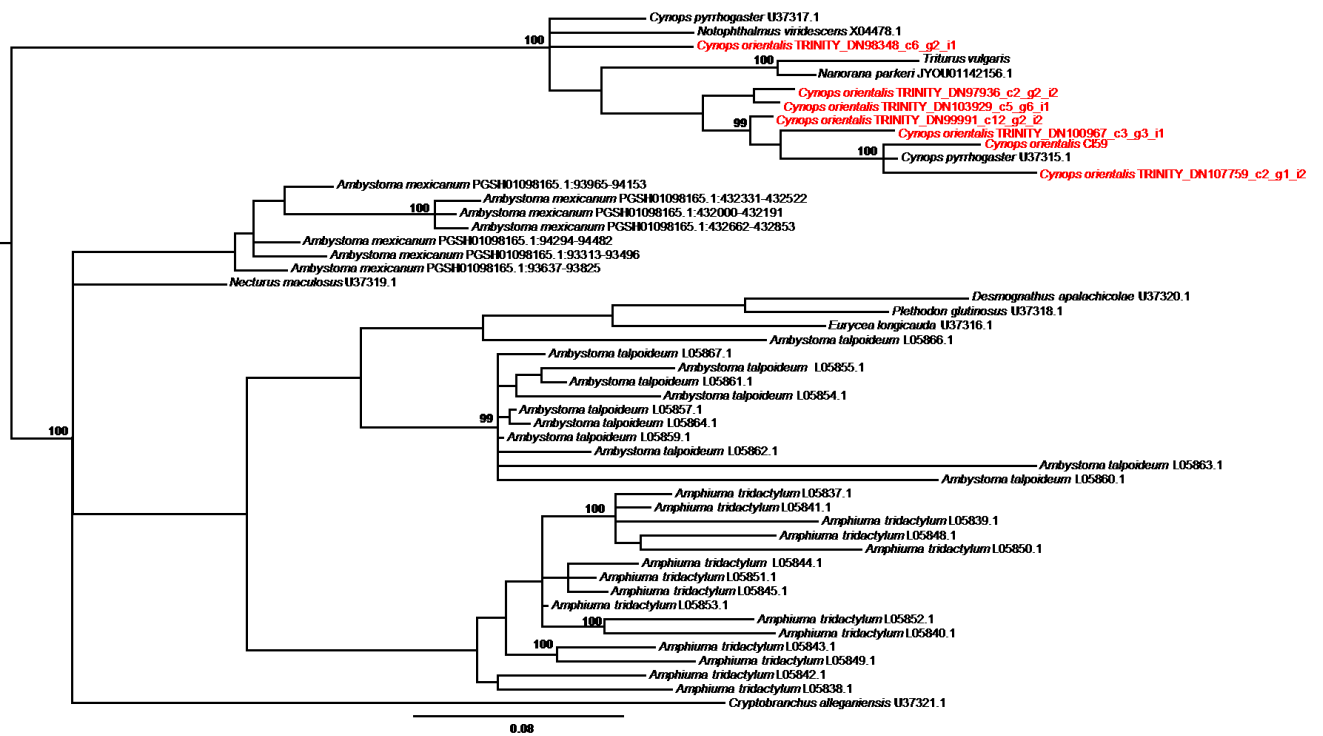


FIGURA 6 Albero filogenetico delle sequenze ottenute su BlastN dall'analisi della sequenza dell'inserto del clone 59p1.

È stato valutato quanto l'elemento analizzato sia presente nel genoma di *C. orientalis* attraverso un dot blot quantitativo, utilizzando concentrazioni scalari sia della sequenza del clone 59p1 che del genoma e come sonda di ibridazione il clone in esame. Dalla FIGURA 7 si evidenzia che il dot della sonda con concentrazione pari a 0,062 ng/μl ha una buona corrispondenza con il dot del genoma con concentrazione pari a 6,25 ng/μl e pertanto la sequenza è presente nel genoma con una percentuale pari all'1%.

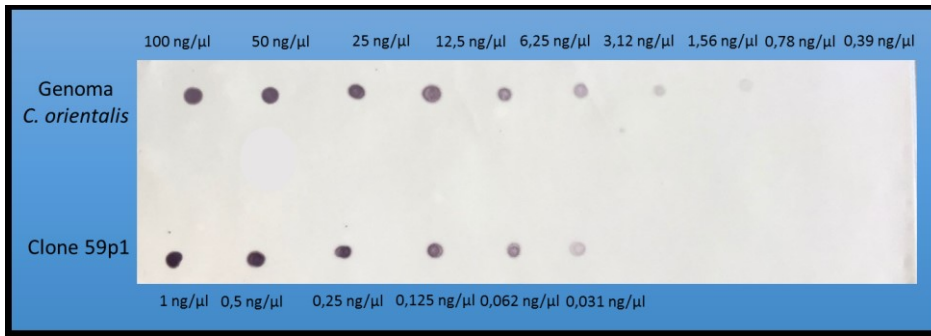


FIGURA 7 Dot blot quantitativo della sequenza inserto del clone 59p1.

Infine è stato eseguito il Southern blot per verificare l'organizzazione della sequenza inserto del clone 59p1 nel genoma di *C. orientalis*. Il pattern d'ibridazione risultante ha mostrato ladder caratteristico di una sequenza organizzata in tandem (Figura 8).

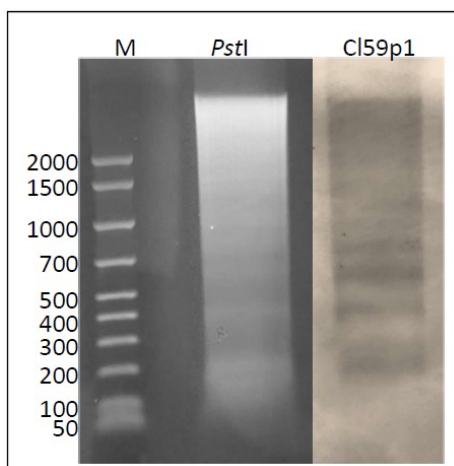


FIGURA 8 Southern blotting su DNA di *C. orientalis* digerito con *PstI* utilizzando come sonda la sequenza inserto del clone 59p1.

Capitolo quarto

DISCUSSIONE

La porzione ripetuta del DNA ha un ruolo importante nelle salamandre che tra gli anfibi hanno mantenuto un genoma di grandi dimensioni durante l'evoluzione. Questo è probabilmente dovuto alla presenza di una grande quantità di sequenze ripetute dovuta probabilmente ad un basso tasso di eliminazione del DNA. Questo studio ha permesso di incrementare le conoscenze in merito, soffermandosi sull'analisi della porzione ripetuta del DNA in *C. orientalis* attraverso la costruzione di una libreria genomica di cui sono state isolate e sequenziate 16 sequenze inserto. Su tre di queste è stata svolta un'analisi più approfondita: due sequenze sono ascrivibili agli elementi trasponibili ed una mostra una organizzazione genomica in tandem del tipo SAT2. Tutte le sequenze identificate in questo lavoro di tesi risultano ancora non caratterizzate in *C. orientalis*.

Gli elementi trasponibili isolati appartengono rispettivamente alla famiglia DIRS e *Ty3/Gypsy*. Nel primo caso si tratta di una sequenza non-LTR, mentre nel secondo di una sequenza LTR. L'analisi di questi risultati permette di descrivere in parte la composizione mista del genoma di *C. orientalis*, andando a verificare quanto sostenuto in altri studi sul genoma delle salamandre (Sun et al., 2012) in cui è stata valutata la presenza e l'abbondanza dei TE. Altresì

importante risulta la caratterizzazione della sequenza di SAT2, dato che le sequenze ripetute in tandem risultano ancora poco caratterizzate nelle salamandre.

4.1. Sequenza Ty3/Gypsy

L'analisi bioinformatica sul clone 1p2 ha permesso di attribuire la sequenza inserto agli elementi trasponibili LTR della famiglia *Ty3/Gypsy*. È presente con una percentuale bassa all'interno del genoma di *C. orientalis*, costituendo solo lo 0,125% del genoma. Tuttavia sequenze appartenenti a questa famiglia sono state valutate essere presenti con valori tra il 7 e il 20% nel genoma delle salamandre (Sun et al., 2012). Questi valori così alti sono possibili solo in genomi di grandi dimensioni, in cui probabilmente la famiglia *Ty3/Gypsy* è stata acquisita durante l'evoluzione tramite trasferimento orizzontale e poi si è diversificata a partire da un TE ancestrale, portando alla formazione di sottofamiglie che risultano attive (Sun & Mueller, 2014). Diversa è la presenza in genomi piccoli, come quello delle piante di *Arabidopsis thaliana* (Marín & Lloréns, 2000), in cui gli LTR non solo sono poco abbondanti ma anche inattivi. Lo stesso vale per i genomi di rettili e uccelli che sono andati incontro ad una riduzione di taglia più rapida a partire dal loro antenato (Sun et al., 2012). Un esempio è lo studio sulla composizione del genoma (2,5 Gb) di *Alligator*

mississippiensis in cui si ha una scarsa presenza di LTR, proprio perché nell'evoluzione sono stati persi (St John et al., 2012).

L'elemento identificato in questo lavoro di tesi è anch'esso attivo perché è stato trovato anche all'interno del trascrittoma di *C. orientalis*, a sostegno del fatto che il gene che codifica per tale proteina è attivo trascrizionalmente. Andando ad analizzare la sequenza di partenza e soffermandosi sulla regione *pol*, si è visto, inoltre, che presenta il sito catalitico DEDD della regione della RNasi-H conservato, suggerendo che anche il prodotto proteico potrebbe essere attivo. Nel caso in cui l'enzima fosse attivo si avrebbe un aumento della ricombinazione, come sostenuto anche in alcuni studi in cui si attribuisce all'enzima un'accelerazione dell'attività dei retrotrasposoni, con conseguente aumento del tasso di trasferimento dei TE nel meccanismo di trasposizione e quindi di accumulo nel genoma (Hellsten et al., 2010; Ustyantsev et al., 2015). Nel caso in cui fosse inattivo, altri autori (Metcalf & Casane, 2013) sostengono che il mantenimento dei TE durante la loro diversificazione sia avvenuto con l'accumulo di mutazioni che ne hanno favorito il loro mantenimento in genomi di grandi dimensioni.

La sequenza dell'RNasi-H è stata utilizzata per identificare, tramite indagine bioinformatica, delle sequenze simili in altri organismi. Le sequenze ottenute sono state allineate ed è stata condotta un'indagine filogenetica. Da

quest'ultima si evince un processo di divergenza evolutiva della sequenza ancestrale dell'RNasi-H in due tipologie rese evidenti tramite i due raggruppamenti di sequenze ottenuti nell'albero: da un lato ci sono le sequenze di sarcopterigi e attinopterigi, ma anche di alcuni echinodermi; dall'altro lato si ha il raggruppamento delle sequenze di *C. orientalis* con *X. tropicalis* e *A. carolinensis*, ma anche dell'echinoderma *S. purpuratus* e di alcuni pesci come *Larimichthys crocea* e *Xiphophorus maculatus*. Da quest'ultimo raggruppamento si individuano due sottogruppi, ove le sequenze degli anfibi, includenti *C. orientalis* e *X. tropicalis*, raggruppano separatamente dalle sequenze di *A. carolinensis*, *S. purpuratus* e pesci. Questo risultato porta a sostenere che una tipologia di RNasi-H è molto diffusa soprattutto nei pesci ma non è stata identificata in *C. orientalis*. La versione presente in quest'ultima mostra una minore diffusione probabilmente perché è andata persa oppure perché è frutto di eventi di trasferimento orizzontale. Inoltre l'identificazione in *Cynops* e in *Xenopus* suggerisce che tale sequenza era già presente nel comune antenato di anuri e caudati.

È stato proposto che dagli elementi della famiglia Ty3/Gypsy sia derivato il dominio SCAN. Esso è lungo circa 80 amminoacidi ed è localizzato nella regione N terminale di molte proteine a dita di zinco, è ricco in residui di leucina e presenta una struttura a α elica. Questo dominio è spesso trovato nelle

proteine aventi il Kruppel-associated box coinvolte nel silenziamento di elementi retrovirali esogeni e endogeni. La sequenza identificata in *Cynops* non presenta il dominio SCAN a differenze di quelle dei cecilidi e degli anuri. Questi dati suggeriscono che tale dominio era presente nell'antenato comune degli anfibi e l'assenza in *Cynops* è attribuibile ad una perdita secondaria.

4.2. Sequenza DIRS

L'analisi bioinformatica condotta sulla sequenza inserto del clone 7p1 ha permesso di attribuire tale sequenza agli elementi trasponibili non-LTR DIRS. L'analisi quantitativa ha permesso di evidenziare che tale sequenza è poco presente nel genoma di *C. orientalis*, in quanto la sua presenza in percentuale è di solo 0,031%. Questo dato rappresenta un valore di confronto con altre specie di salamandre che presentano una percentuale di sequenze DIRS con valori tra 2 e 5.7% del genoma (Sun et al., 2012). Tuttavia è noto che i DIRS siano più presenti nel genoma di pesci attinopterigi, sono mantenuti anche negli anfibi e in alcuni rettili (Kordis, 2010), ma vengono persi in uccelli e mammiferi (Chalopin et al., 2015; Shao et al., 2019).

Analizzando la sequenza del clone si è visto che l'elemento DIRS presenta una regione che ha similarità con una sequenza di *X. tropicalis* appartenente alla superfamiglia di DNA breaking-joining enzymes. Queste proteine includono le

topoisomerasi e le tirosina-ricombinasi, quest'ultimi enzimi caratteristici della regione C-terminale dei DIRS.

La sequenza inserto è presente anche nel trascrittoma di *C. orientalis*, quindi è una regione che viene trascritta ma di cui non è nota l'attività a livello proteico.

4.3. Sequenza SAT2

L'analisi bioinformatica della sequenza inserto del clone 59p1 ha permesso di evidenziarne l'appartenenza al DNA satellite SAT2. Si tratta di una sequenza ripetuta in tandem, come risulta dal Southern blot che mostra il caratteristico ladder delle sequenze che presentano questo tipo di organizzazione. Inoltre è presente nel genoma di *C. orientalis* con una percentuale pari all' 1%.

SAT2 è stato trovato nel trascrittoma per cui si potrebbe e questo suggerisce che abbia dei promotori attivi nei siti vicini che ne permettono la trascrizione. In anfibi e negli uccelli questo tipo di DNA viene trascritto a livello dei cromosomi a spazzola in *non-coding* RNA, che è coinvolto nella formazione e nel mantenimento dell' eterocromatina centromerica e telomerica, con impatto sulla stabilità e integrità genomica (Biscotti et al., 2015). Poco nota è l'attività nei vari processi di gametogenesi (Trofimova & Krasikova, 2016). Altri studi, incentrati proprio su SAT2, sottolineano la presenza di queste sequenze nel trascrittoma come un meccanismo chiave per favorire la loro propagazione nel

genoma, attraverso un meccanismo simile alla retrotrasposizione e che permette anche l'integrazione di satDNA in sequenze intersperse (Zhang & Epstein, 1996).

L'analisi filogenetica ha permesso di evidenziare che la sequenza isolata in *C. orientalis* sia presente in molte salamandre. In particolare è bene sottolineare una possibile implicazione nell'evoluzione della sequenza stessa dato che si ha una separazione netta tra il gruppo che include varie specie di *Ambystoma* (Green et al., 1993) e quello che include varie sequenze appartenenti a due specie del genere *Cynops* che raggruppano con *Triturus vulgaris* (Barsacchi-Pilone et al., 1986) e *Notophthalmus viridescens* (Epstein et al., 1986). Infatti già altri studi avevano evidenziato la presenza di SAT2 all'interno di tre famiglie di salamandre: *Ambystoma talpoideum*, *Amphiuma tridactylum* e *Notophthalmus viridescens*. Anche in quel caso *A. talpoideum* e *A. tridactylum* presentavano una similarità maggiore rispetto a *N. viridescens* (Green et al., 1993). Ma in questo studio si evince che le sequenze di quest'ultimo hanno una maggiore similarità con quelle identificate in *Cynops* e soprattutto che in questo gruppo è inclusa anche la sequenza di *Nanorana parkeri*. Questo dimostra che la sequenza era presente già dall'antenato comune di anuri e caudata, diversamente da quanto sostenuto negli studi che avevano sottolineato l'assenza di SAT2 in *X. laevis*, appartenente agli anuri (Zhang & Epstein,

1996). Inoltre, l'analisi condotta sui genomi disponibili di diverse specie di cecilidi consente di confermare l'assenza della sequenza nell'antenato comune degli anfibi, come sostenuto nello stesso studio in cui si considerava l'assenza di SAT2 in *Typhlonectes natans*, appartenente ai Gymnophiona.

Bibliografia

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. [https://doi.org/https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/https://doi.org/10.1016/S0022-2836(05)80360-2)
- Barsacchi-Pilone, G., Batistoni, R., Andronico, F., Vitelli, L., & Nardi, I. (1986). Heterochromatic DNA in Triturus (Amphibia, Urodela). I. A satellite DNA component of the pericentric C-bands. *Chromosoma*, *93*(5), 435–446. <https://doi.org/10.1007/BF00285826>
- Benachenhou, F., Sperber, G. O., Bongcam-Rudloff, E., Andersson, G., Boeke, J. D., & Blomberg, J. (2013). Conserved structure and inferred evolutionary history of long terminal repeats (LTRs). *Mobile DNA*, *4*(1). <https://doi.org/10.1186/1759-8753-4-5>
- Biscotti, M. A., Carducci, F., Barucca, M., Gerdol, M., Pallavicini, A., Scharl, M., Canapa, A., & Adolfini, M. C. (2020). The transcriptome of the newt *Cynops orientalis* provides new insights into evolution and function of sexual gene networks in sarcopterygians. *Scientific Reports*, *10*(1), 1–14. <https://doi.org/10.1038/s41598-020-62408-x>
- Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. (Pat. (2015). Repetitive DNA in eukaryotic genomes. *Chromosome Research*, *23*(3), 415–420. <https://doi.org/10.1007/s10577-015-9499-z>
- Bode, H. R. (2003). Head regeneration in Hydra. *Developmental Dynamics*, *226*(2), 225–236. <https://doi.org/10.1002/dvdy.10225>
- Bonett, R. M., Chippindale, P. T., Moler, P. E., van Devender, R. W., & Wake, D. B. (2009). Evolution of gigantism in amphiumid salamanders. *PLoS ONE*, *4*(5). <https://doi.org/10.1371/journal.pone.0005615>
- Buckley, D. (2012). Evolution of Viviparity in Salamanders (Amphibia, Caudata). In *eLS*. <https://doi.org/10.1002/9780470015902.a0022851>
- Canapa, A., Barucca, M., Biscotti, M. A., Forconi, M., & Olmo, E. (2016). Transposons, genome size, and evolutionary insights in animals. *Cytogenetic and Genome Research*, *147*(4), 217–239. <https://doi.org/10.1159/000444429>
- Cappello, J., Handelsman, K., & Lodish, H. F. (1985). Sequence of Dictyostelium DIRS-1: An apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. *Cell*, *43*(1), 105–115.

[https://doi.org/10.1016/0092-8674\(85\)90016-9](https://doi.org/10.1016/0092-8674(85)90016-9)

- Carducci, F., Biscotti, M. A., Barucca, M., & Canapa, A. (2019). Transposable elements in vertebrates: species evolution and environmental adaptation. *European Zoological Journal*, 86(1), 497–503. <https://doi.org/10.1080/24750263.2019.1695967>
- Chalopin, D., Naville, M., Plard, F., Galiana, D., & Volff, J. N. (2015). Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biology and Evolution*, 7(2), 567–580. <https://doi.org/10.1093/gbe/evv005>
- Charlesworth, B., Sniegowski, P., & Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, 371, 215–220. <https://www.nature.com/articles/371215a0.pdf>
- Cooper, J. L., & Henikoff, S. (2004). Adaptive evolution of the histone fold domain in centromeric histones. *Molecular Biology and Evolution*, 21(9), 1712–1718. <https://doi.org/10.1093/molbev/msh179>
- Denoël, M., Joly, P., & Whiteman, H. H. (2005). Evolutionary ecology of facultative paedomorphosis in newts and salamanders. *Biological Reviews of the Cambridge Philosophical Society*, 80(4), 663–671. <https://doi.org/10.1017/S1464793105006858>
- Dodonova, S. O., Prinz, S., Bilanchone, V., Sandmeyer, S., & Briggs, J. A. G. (2019). Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses. *Proceedings of the National Academy of Sciences of the United States of America*, 116(20), 10048–10057. <https://doi.org/10.1073/pnas.1900931116>
- Epstein, L. M., Mahon, K. A., & Gall, J. G. (1986). Transcription of a satellite DNA in the newt. *Journal of Cell Biology*, 103(4), 1137–1144. <https://doi.org/10.1083/jcb.103.4.1137>
- Faivovich, N., Frost, D. R., Grant, T., Haddad, L. I. O. F. B., Bain, R. H., Haas, A., Channing, A., Wilkinson, M., Sa, R. O. D. E., Donnellan, S. C., Raxworthy, C. J., Campbell, J. A., Blotto, B. L., Moler, P., Drewes, R. C., Nussbaum, R. A., & Lynch, J. D. (2006). The amphibian tree of life. In *Bulletin of the American Museum of Natural History* (Issue 297).
- Finnegan, D. (1989). Eukaryotic transposable elements and genome evolution. *Trends in Genetics: TIG*, 5(103–107), 103–107. [https://doi.org/10.1016/0168-9525\(89\)90039-5](https://doi.org/10.1016/0168-9525(89)90039-5)

- Gardner, J. D., Laurin, M., & Organ, C. L. (2020). The relationship between genome size and metabolic rate in extant vertebrates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1793). <https://doi.org/10.1098/rstb.2019.0146>
- Garrido-Ramos, M. A. (2017). Satellite DNA: An evolving topic. *Genes*, 8(9). <https://doi.org/10.3390/genes8090230>
- Gentles, A. J., Kohany, O., & Jurka, J. (2005). Evolutionary diversity and potential recombinogenic role of integration targets of non-LTR retrotransposons. *Molecular Biology and Evolution*, 22(10), 1983–1991. <https://doi.org/10.1093/molbev/msi188>
- Grandi, F. C., Rosser, J. M., & An, W. (2013). LINE-1-derived poly(A) microsatellites undergo rapid shortening and create somatic and germline mosaicism in mice. *Molecular Biology and Evolution*, 30(3), 503–512. <https://doi.org/10.1093/molbev/mss251>
- Grau, J. H., Poustka, A. J., Meixner, M., & Plötner, J. (2014). LTR retroelements are intrinsic components of transcriptional networks in frogs. *BMC Genomics*, 15(1), 1–16. <https://doi.org/10.1186/1471-2164-15-626>
- Green, B., Pabdn-pen, L. M., Graham, T. A., Peach, S. E., Coats, S. R., & Epstein, L. A. N. (1993). Conserved sequence and functional domains in satellite 2 from three families of salamanders. *Molecular Biology and Evolution*, 10(4), 732–750. <https://doi.org/10.1093/oxfordjournals.molbev.a040041>
- Gregory, T.R. (2001). Coincidence, coevolution, or causation? DNA content, cellsize, and the C-value enigma. *Biological Reviews*, 76(1), 65–101. <https://doi.org/10.1111/j.1469-185X.2000.tb00059.x>
- Gregory, T.R. (2020). *Animal Genome Size Database*. <https://www.genomesize.com>.
- Hall, B. K. (2003). Evo-Devo: Evolutionary developmental mechanisms. *International Journal of Developmental Biology*, 47(7–8), 491–495. <https://doi.org/10.1387/ijdb.14756324>
- Hanken, J., & Wake, D. B. (1993). Miniaturization of body size: Organismal consequences and evolutionary significance. *Annual Review of Ecology and Systematics*, 24, 501–519. <https://doi.org/10.1146/annurev.es.24.110193.002441>

- Hanken, James. (1983). Miniaturization and its effects on cranial morphology in plethodontid salamanders, genus *Thorius* (Amphibia, Plethodontidae): II. The fate of the brain and sense organs and their role in skull morphogenesis and evolution. *Journal of Morphology*, 177(3), 255–268. <https://doi.org/10.1002/jmor.1051770304>
- Hedges, D. J., & Deininger, P. L. (2007). Inviting Instability: Transposable elements, Double-strand breaks, and the Maintenance of Genome Integrity. *Mutat Res*, 616(1-2), 46–59.
- Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N. H., Shu, S., Taher, L., Blitz, I. L., Blumberg, B., Dichmann, D. S., Dubchak, L., Amaya, E., Detter, J. C., Fletcher, R., Gerhard, D. S., Goodstein, D., Rokhsar, D. S. (2010). The genome of the western clawed frog *xenopus tropicalis*. *Science*, 328(5978), 633–636. <https://doi.org/10.1126/science.1183670>
- Huelsenbeck, J. P., Larget, B., Miller, R. E., & Ronquist, F. (2002). Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology*, 51(5), 673–688. <https://doi.org/10.1080/10635150290102366>
- Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, 110(1–4), 462–467. <https://doi.org/10.1159/000084979>
- Kordis, D. (2010). Transposable elements in reptilian and avian (Sauropsida) genomes. *Cytogenetic and Genome Research*, 127(2–4), 94–111. <https://doi.org/10.1159/000294999>
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
- Levy, D. L., & Heald, R. (2016). Biological scaling problems and solutions in amphibians. *Cold Spring Harbor Perspectives in Biology*, 8(1), 1–16. <https://doi.org/10.1101/cshperspect.a019166>
- Lynch Michael. (2007). The Origins of Genome Architecture. In *Journal of Heredity* (Vol. 98, Issue 6, pp. 633–634). <https://doi.org/10.1093/jhered/esm073>

- Makałowski, W., Gotea, V., Pande, A., & Makałowska, I. (2019). Transposable elements: Classification, identification, and their use as a tool for comparative genomics. In *Methods in Molecular Biology* (Vol. 1910, pp. 177–207). https://doi.org/10.1007/978-1-4939-9074-0_6
- Malik, H. S., Henikoff, S., & Eickbush, T. H. (2000). Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Research*, *10*(9), 1307–1318. <https://doi.org/10.1101/gr.145000>
- Marín, I., & Lloréns, C. (2000). Ty3/Gypsy retrotransposons: Description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. *Molecular Biology and Evolution*, *17*(7), 1040–1049. <https://doi.org/10.1093/oxfordjournals.molbev.a026385>
- Metcalf, C., & Casane, D. (2013). Accommodating the load: The transposable element content of very large genomes. *Mobile Genetic Elements*, *3*(2), e24775. <https://doi.org/10.4161/mge.24775>
- Montermini, L., Andermann, E., Labuda, M., Richter, A., Pandolfo, M., Cavalcanti, F., Pianese, L., Iodice, L., Farina, G., Monticelli, A., Turano, M., Filla, A., De Michele, G., & Coccozza, S. (1997). The Friedreich ataxia GAA triplet repeat: Premutation and normal alleles. *Human Molecular Genetics*, *6*(8), 1261–1266. <https://doi.org/10.1093/hmg/6.8.1261>
- Muszewska, A., Steczkiewicz, K., & Ginalski, K. (2013). DIRS and Ngara Retrotransposons in Fungi. *PLoS ONE*, *8*(9), 1–10. <https://doi.org/10.1371/journal.pone.0076319>
- Naville, M., Warren, I. A., Haftek-Terreau, Z., Chalopin, D., Brunet, F., Levin, P., Galiana, D., & Volff, J. N. (2016). Not so bad after all: Retroviruses and long terminal repeat retrotransposons as a source of new genes in vertebrates. *Clinical Microbiology and Infection*, *22*(4), 312–323. <https://doi.org/10.1016/j.cmi.2016.02.001>
- Paço, A., Freitas, R., & Vieira-Da-Silva, A. (2019). Conversion of DNA sequences: From a transposable element to a tandem repeat or to a gene. *Genes*, *10*(12). <https://doi.org/10.3390/genes10121014>
- Pagni, M., & Junier, T. (1999). *Dotlet JS beta program*. https://myhits.sib.swiss/util/dotlet/doc/dotlet_help.html
- Ribet, D., Harper, F., Esnault, C., Pierron, G., & Heidmann, T. (2008). The GLN Family of Murine Endogenous Retroviruses Contains an Element Competent for Infectious Viral Particle Formation. *Journal of Virology*,

82(9), 4413–4419. <https://doi.org/10.1128/jvi.02141-07>

- Richard, G.-F., Kerrest, A., & Dujon, B. (2008). Comparative Genomics and Molecular Dynamics of DNA Repeats in Eukaryotes. *Microbiology and Molecular Biology Reviews*, 72(4), 686–727. <https://doi.org/10.1128/membr.00011-08>
- Rink, J. C. (2013). Stem cell systems and regeneration in planaria. *Development Genes and Evolution*, 223(1–2), 67–84. <https://doi.org/10.1007/s00427-012-0426-4>
- Rodríguez, F., Oliver, J. L., Marín, A., & Medina, J. R. (1990). The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology*, 142(4), 485–501. [https://doi.org/https://doi.org/10.1016/S0022-5193\(05\)80104-3](https://doi.org/https://doi.org/10.1016/S0022-5193(05)80104-3)
- Rubin, B. P., Brockes, J., Galliot, B., Grossniklaus, U., Lobo, D., Mainardi, M., Mirouze, M., Prochiantz, A., & Steger, A. (2015). A dynamic architecture of life. *F1000Research*, 4(0), 1–17. <https://doi.org/10.12688/f1000research.7315.1>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
- Sessions, S. K., & Larson, A. (1987). Developmental Correlates of Genome Size in Plethodontid Salamanders and Their Implications for Genome Evolution. *Evolution*, 41(6), 1239. <https://doi.org/10.2307/2409090>
- Shao, F., Han, M., & Peng, Z. (2019). Evolution and diversity of transposable elements in fish genomes. *Scientific Reports*, 9(1), 1–8. <https://doi.org/10.1038/s41598-019-51888-1>
- Sotero-Caio, C. G., Platt, R. N., Suh, A., & Ray, D. A. (2017). Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biology and Evolution*, 9(1), 161–177. <https://doi.org/10.1093/gbe/evw264>
- St John, J. A., Braun, E. L., Isberg, S. R., Miles, L. G., Chong, A. Y., Gongora, J., Dalzell, P., Moran, C., Bed’Hom, B., Abzhanov, A., Burgess, S. C., Cooksey, A. M., Castoe, T. A., Crawford, N. G., Densmore, L. D., Drew, J. C., Edwards, S. V., Faircloth, B. C., Fujita, M. K., ... Ray, D. A. (2012). Sequencing three crocodylian genomes to illuminate the evolution of archosaurs and amniotes. *Genome Biology*, 13(1), 1–12.

<https://doi.org/10.1186/gb-2012-13-1-415>

- Sun, C., López Arriaza, J. R., & Mueller, R. L. (2012). Slow DNA loss in the gigantic genomes of salamanders. *Genome Biology and Evolution*, 4(12), 1340–1348. <https://doi.org/10.1093/gbe/evs103>
- Sun, C., & Mueller, R. L. (2014). Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders. *Genome Biology and Evolution*, 6(7), 1818–1829. <https://doi.org/10.1093/gbe/evu143>
- Sun, C., Shepard, D. B., Chong, R. A., Arriaza, J. L., Hall, K., Castoe, T. A., Feschotte, C., Pollock, D. D., & Mueller, R. L. (2012). LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biology and Evolution*, 4(2), 168–183. <https://doi.org/10.1093/gbe/evr139>
- Tang, J., Yu, Y., Zheng, H., Yin, L., Sun, M., Wang, W., Cui, J., Liu, W., Xie, X., & Chen, F. (2017). ITRAQ-based quantitative proteomic analysis of *Cynops orientalis* limb regeneration. *BMC Genomics*, 18(1), 1–12. <https://doi.org/10.1186/s12864-017-4125-4>
- Tollis, M., & Boissinot, S. (2013). Lizards and LINEs: Selection and demography affect the fate of L1 retrotransposons in the genome of the green anole (*Anolis carolinensis*). *Genome Biology and Evolution*, 5(9), 1754–1768. <https://doi.org/10.1093/gbe/evt133>
- Trofimova, I., & Krasikova, A. (2016). Transcription of highly repetitive tandemly organized DNA in amphibians and birds: A historical overview and modern concepts. *RNA Biology*, 13(12), 1246–1257. <https://doi.org/10.1080/15476286.2016.1240142>
- Ugarkovic, D. (2005). Functional elements residing within satellite DNAs. *EMBO Reports*, 6(11), 1035–1039. <https://doi.org/10.1038/sj.embor.7400558>
- Ustyantsev, K., Novikova, O., Blinov, A., & Smyshlyaev, G. (2015). Convergent evolution of ribonuclease H in LTR retrotransposons and retroviruses. *Molecular Biology and Evolution*, 32(5), 1197–1207. <https://doi.org/10.1093/molbev/msv008>
- Vignali, R., & Nardi, I. (1996). Unusual features of the urodele genome: Do they have a role in evolution and development? *International Journal of Developmental Biology*, 40(4), 637–643. <https://doi.org/10.1387/ijdb.8877435>

- Wake, D. B., & Koo, M. S. (2018). Amphibians. *Current Biology*, 28(21), R1237–R1241. <https://doi.org/10.1016/j.cub.2018.09.028>
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2008). A universal classification of eukaryotic transposable elements implemented in Repbase JO. *Nature Reviews Genetics SP*, 05(01). <https://doi.org/10.1038/nrg2165-c2> ID
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2009). Reply: A unified classification system for eukaryotic transposable elements should reflect their phylogeny. *Nature Reviews Genetics*, 10(4), 276. <https://doi.org/10.1038/nrg2165-c4>
- Widlund, H. R., Kuduvalli, P. N., Bengtsson, M., Cao, H., Tullius, T. D., & Kubista, M. (1999). Nucleosome structural features and intrinsic properties of the TATAAACGCC repeat sequence. *Journal of Biological Chemistry*, 274(45), 31847–31852. <https://doi.org/10.1074/jbc.274.45.31847>
- Wu, Y. (2020). *AmphibiaWeb 2008 Cynops orientalis: Oriental Fire-bellied Newt*. [Http://Amphibiaweb.Org/Species/4244](http://Amphibiaweb.Org/Species/4244).
- Zhang, Y., & Epstein, L. M. (1996). Cloning and characterization of extended hammerheads from a diverse set of caudate amphibians. *Gene*, 172(2), 183–190. [https://doi.org/10.1016/0378-1119\(96\)00126-6](https://doi.org/10.1016/0378-1119(96)00126-6)

