



Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN BIOLOGIA EVOLUZIONISTICA E AMBIENTALE

CICLO XXI°

CURRICULUM: GENETICA

COORDINATORE Prof. Guido Barbujani

CAMBIAMENTI TEMPORALI NELLA DIVERSITÀ GENETICA DEL TONNO ROSSO (*Thunnus thynnus*) DEL MEDITERRANEO

Settore Scientifico Disciplinare BIO/18

Dottorando

Dott. Riccioni Giulia

(firma)

Tutore

Prof. Barbujani Guido

(firma)

Anni 2006/2008

INDICE

1. INTRODUZIONE.....	4
1.1. Ecologia del <i>Thunnus thynnus</i> e problematiche relative alla sua conservazione.....	4
1.1.1. Bio-ecologia.....	6
1.1.2. Analisi Genetiche.....	11
1.1.3. Sfruttamento del tonno rosso nel XX° secolo.....	13
1.2. Cos'è la genetica della conservazione?.....	17
1.2.1. Diagnosticare i problemi genetici.....	18
1.3. Alberi genealogici, teoria della coalescenza e analisi dei polimorfismi genetici.....	21
1.3.1. Modello standard di coalescenza.....	23
1.4. La statistica Bayesiana in genetica.....	26
1.4.1. Principi di inferenza Bayesiani.....	27
1.4.2. Metodi Monte Carlo Markov Chain.....	28
1.4.3. Integrazione Monte Carlo.....	29
1.4.4. Algoritmo Metropolis-Hastings.....	29
1.4.5. Applicazioni alla genetica di popolazioni.....	30
2. SCOPO DELLO STUDIO.....	32
3. MATERIALI E METODI.....	33
3.1. Stima delle statistiche descrittive.....	33
3.2. Struttura di popolazione.....	38

3.3. Stime demografiche da confronti temporali tra campioni.....	38
3.4. Stime demografiche da singoli campioni.....	40
3.5. Verifica degli effetti di crisi demografiche.....	42
4. RISULTATI.....	44
4.1. Analisi delle statistiche descrittive.....	44
4.2. Analisi di struttura di popolazione.....	47
4.3. Inferenza stime demografiche.....	48
4.4. Identificazione dei segnali di crisi demografica.....	50
5. DISCUSSIONE.....	53
6. BIBLIOGRAFIA.....	58

1.INTRODUZIONE

1.1. Ecologia del *Thunnus thynnus* e problematiche relative alla sua conservazione

Il tonno rosso (*Thunnus thynnus*, in inglese *Atlantic Bluefin tuna*) è il più grande tra le specie della famiglia degli Scombridi ed uno dei grandi predatori dell'ecosistema pelagico. *T. thynnus* possiede adattamenti morfo-fisiologici altamente specializzati che gli consentono di sfruttare un ampio range di habitat pelagici, dalle acque tropicali a quelle sub-artiche dell'oceano Atlantico. La specie presenta un'ampia distribuzione nell'Atlantico settentrionale e centrale (Figura 1). E' una specie altamente migratoria e sono state ampiamente documentate traversate transatlantiche (Block *et al.* 2005, Fromentin and Powers 2005, Rooker *et al.* 2008). Questi spostamenti degli individui su un'ampia scala spaziale tra le zone di riproduzione e quelle di alimentazione determinano un'incoerenza tra limiti delle popolazioni biologiche e limiti di definizione degli stock (cioè le unità funzionali di gestione della pesca). Per questo una cooperazione internazionale coordinata basata sull'acquisizione di dati scientifici e che mira ad una gestione ecocompatibile della risorsa è necessaria per la gestione effettiva del *Thunnus thynnus* e in generale delle specie di pesci grandi migratori. In particolare, le decisioni gestionali adottate su scala spaziale e temporale ampie dipendono in modo rilevante dall'acquisizione di conoscenze relative alla dinamica e alla storia demografica delle popolazioni e alla identificazione di unità biologiche indipendenti (le popolazioni in senso strettamente biologico) all'interno delle unità di gestione o stock.

Attualmente per il tonno rosso si riconoscono due zone principali di riproduzione: il mar Mediterraneo per lo stock orientale ed il Golfo del Messico per lo stock occidentale (Figura 1). Sebbene siano state descritte migrazioni transatlantiche, gli esperimenti condotti con marcature elettroniche degli individui hanno rivelato un alto grado di fedeltà alle zone di nascita e questo ha portato la Commissione Internazionale per la Conservazione del Tonno Atlantico (ICCAT) ad identificare due stock (orientale ed occidentale) i quali vengono considerati per la valutazione dei parametri bio-ecologici e demografici indipendenti. L'elevato numero di individui che dopo aver visitato le aree di alimentazione ritorna alle zone di riproduzione e la presenza di un lieve ma significativo differenziamento genetico (per una revisione dei contributi scientifici relativi vedi Fromentin and Powers 2005; Rooker *et al.* 2007) rappresentano prove coerenti a favore dell'indipendenza dei due stock.

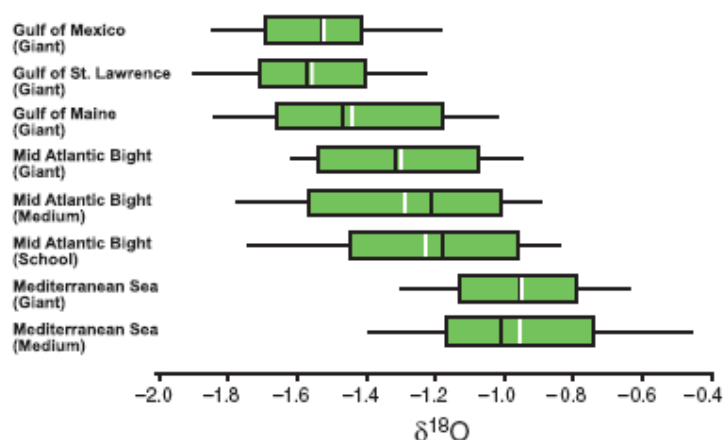


Figura 1. Distribuzione spaziale del tonno rosso (area in grigio) e principali rotte migratorie (freccie nere) sono state dedotte sia da dati di pesca storici e attuali che da informazioni ottenute da tag elettroniche e tradizionali. La linea verticale tratteggiata rappresenta la delimitazione dello stock nelle due unità correnti di gestione dell'ICCAT. I siti di riproduzione (aree punteggiate grigie) sono localizzate nel mar Mediterraneo e nel Golfo del Messico (Fromentin and Powers 2005).

Tuttavia, recenti studi basati sull'analisi della composizione di microelementi negli otoliti di individui dei due stock hanno definito che una parte significativa degli individui di tonno rosso pescati nelle zone di alimentazione delle coste americane è formata da individui verosimilmente nati nel Mediterraneo (Rooker *et al.* 2008; Figura 2).

Figura 2.

Box plot che mostrano i valori di $\delta^{18}\text{O}$ rilevati negli individui giovani (<60kg), medi (60-140 kg) e giganti (>140 kg) di tonno rosso dell'Atlantico dalle zone di riproduzione (mar Mediterraneo, Golfo del Messico), dalle zone di alimentazione (Golfo di S. Lorenzo, Golfo del Maine, Baia del medio Atlantico). I box mostrano il 25° e 75° percentile e la barra di errore rappresenta il 10° e 90° percentile. La mediana è indicata come una barra nera e la media come una barra bianca. Date di campionamento: mar Mediterraneo (2003-2007), Golfo del Messico (2004,2007), Golfo di S. Lorenzo (2006-2007), Golfo del Maine (1996,1998), Baia del medio Atlantico (1997-200).



La controversia sull'indipendenza delle popolazioni dei due stock è di particolare interesse per la gestione in ragione delle rilevanti differenze demografiche e riproduttive che mostrano. Infatti la biomassa riproduttiva dello stock occidentale è assai minore ed è drasticamente crollata negli ultimi decenni nonostante le severe restrizioni alla pesca applicate su questo stock da oltre 20 anni. Al contrario, lo stock orientale è significativamente più grande e più stabile, anche se le numerose incertezze sulle statistiche di pesca del tonno rosso nel Mediterraneo alimentano fortemente il dibattito sulle condizioni di questo stock (ICCAT 2007).

Nell'ultimo decennio, i ricercatori hanno contribuito in modo significativo alle conoscenze dei tratti bio-ecologici e demografici del tonno rosso relativamente ad entrambi gli stock arrivando in particolare ad acquisire dati pertinenti sui parametri biologici di base (come età di maturità, fecondità, alimentazione, crescita) (vedi Rooker *et al.* 2007 e lavori citati) che sono determinanti per la valutazione delle condizioni di uno stock. Avanzamenti significativi delle conoscenze sono stati acquisiti anche grazie all'applicazione di tecnologie innovative per lo studio delle popolazioni, come l'analisi della struttura genetica con loci marcatori ipervariabili o microsatelliti o, come illustrato in precedenza, l'analisi della composizione di microelementi (microchimica) degli otoliti. Lo sviluppo tecnologico offre quindi nuove prospettive per lo studio dell'ecologia e dell'evoluzione delle popolazioni di *T. thynnus*.

1.1.1. Bio-ecologia

La riproduzione del tonno rosso nelle due aree (Mediterraneo e Golfo del Messico) appare fortemente correlata alle specifiche condizioni oceanografiche che rappresentano un compromesso tra le necessità delle larve e le limitazioni fisiologiche degli adulti. Il periodo di riproduzione sembra essere legato prevalentemente alla temperatura superficiale dell'acqua. Poiché la temperatura dell'acqua nel Golfo del Messico supera i 24° C (soglia stimata per la riproduzione) ad inizio primavera, in questa area *T. thynnus* inizia l'accoppiamento in anticipo rispetto al Mediterraneo (aprile nel Golfo del Messico, maggio nel Mediterraneo).

Nel Mediterraneo le larve di tonno rosso sono presenti pressoché in tutto il bacino, ad eccezione del mar Ligure, nord Adriatico e mare di Alboran. Tuttavia, in ragione di una maggiore frequenza di rinvenimento delle larve, sono state identificate delle principali zone di riproduzione (spawning areas) corrispondenti alle acque del canale di Sicilia (vicino Capo Passero e canale di Malta) all'arcipelago delle Baleari (canale di Maiorca e a sud di Minorca, Figura 3).

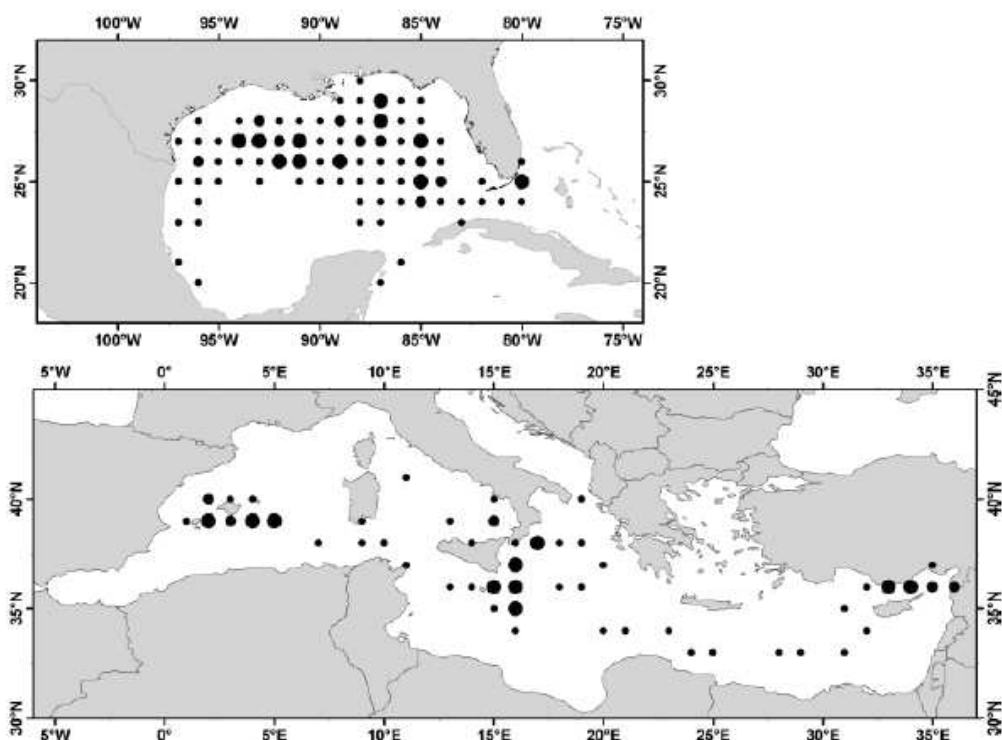


Figura 3. Distribuzione delle larve di tonno rosso dell'atlantico (*Thunnus thynnus*) nel Golfo del Messico e nel mar Mediterraneo ottenuta dalla letteratura (Nishida *et al.* 1998, Garcia *et al.* 2003, Oray and Karakulak 2005) e da dati non pubblicati (dati SEAMAP) forniti da S. Turner e W. Richards. I cerchi piccoli rappresentano catture di 1-4 larve, quelli medi 5-10 larve, i cerchi più larghi 10 o più larve. (Rooker *et al.* 2007).

Inoltre, studi recenti hanno indicato la presenza di una ulteriore area di riproduzione nel Mediterraneo orientale, corrispondente alle acque a nord di Cipro (lungo le coste della Turchia; (Oray and Karakulak 2005). Il periodo di accoppiamento stimato, basandosi sulla condizione delle gonadi, avviene tra maggio e giugno nel Mediterraneo orientale e tra giugno e luglio nel Mediterraneo centro-occidentale. Anche in questo caso le differenze nel periodo di riproduzione tra gli individui abitanti nelle due aree è probabilmente legato alla variazione della temperatura superficiale dell'acqua che raggiunge 24°C già verso metà-fine maggio nel Mediterraneo orientale.

Tuttavia, esistono evidenze che i demi (sottopopolazioni o gruppi di individui identificabili a livello geografico) possono avere dinamiche ecologiche e riproduttive indipendenti verosimilmente legate a differenze ambientali. Durante il periodo riproduttivo sono stati infatti individuati esemplari sessualmente maturi di *T. thynnus* fuori dalle aree di riproduzione. Block *et al.* (2005), mediante il monitoraggio continuo degli individui con marcatori satellitari ha osservato che diversi adulti (>200cm) rimangono nella piattaforma

continentale nord americana e nelle acque dell'Atlantico centrale durante la primavera e l'estate. Nello specifico, sulla base di questi dati i ricercatori hanno formulato tre ipotesi: 1) l'accoppiamento avviene anche in altre regioni oltre al Golfo del Messico e al Mediterraneo; 2) una frazione degli adulti non sono maturi dal punto di vista riproduttivo; 3) il tonno rosso non si riproduce tutti gli anni, ma a volte salta l'accoppiamento. Campionamenti effettuati nelle zone centrali del nord Atlantico non hanno prodotto esiti positivi e larve di tonno rosso non sono state individuate neanche da collezioni di ittioplankton. Al contrario avvistamenti di tonni con taglia oltre quella di prima maturità sessuale, ma di età inferiore, sono stati ritrovati nel Golfo del Messico da Block *et al.* (2005). Perciò tonni delle stesse dimensioni individuati al di fuori dell'area riproduttiva nel periodo di riproduzione potrebbero non essere ancora maturi. L'ultima ipotesi necessita ulteriori approfondimenti, anche se il monitoraggio attraverso marcature elettroniche e tradizionali (Block *et al.* 2005, Teo *et al.* 2007) sembra escludere la possibilità di un 'salto' dell'accoppiamento.

Il numero medio di uova per grammo di peso corporeo stimato nel *T. thynnus* è >90 oociti/g, molto maggiore di quello stimato per altri tonni del genere *Thunnus*, in cui la fecondità individuale stimata è meno di 70 oociti/g. Per quanto riguarda la taglia e l'età di prima maturità i dati a disposizione sono limitati ed emergono grosse differenze tra lo stock dell'est e dell'ovest Atlantico. Nello studio di Corriero *et al.* (Corriero *et al.* 2005), il 50% degli esemplari di tonno rosso del Mediterraneo raggiungono la maturità sessuale ad una lunghezza alla forca di 104 cm (3-4 anni di età) e il 100% sono maturi a 130 cm (5 anni di età). Questi dati sono in contrasto con quelli ottenuti mediante marcature elettroniche. Sulla base dei movimenti migratori differenziali in individui di differente taglia (esemplari adolescenti di *T. thynnus* segnalati al largo del nord Carolina e destinati a ritornare nel Mediterraneo durante la stagione degli accoppiamenti non rientravano fino a quando non raggiungevano una lunghezza maggiore di 205 cm), Block *et al.* (2005) hanno ipotizzato che i tonni del Mediterraneo si riproducono solo ad un'età assai più tardiva (9-10 anni). Le analisi istologiche degli ovari di femmine di tonno rosso dell'Atlantico occidentale mostrano una maturazione ritardata rispetto a quelle del Mediterraneo e sembra improbabile che gli individui raggiungano la maturità prima degli 8 anni. Studi seguenti hanno confermato che gli esemplari dello stock occidentale raggiungono la maturità a circa dieci anni di età e 200 cm. Appare quindi chiaro che lo stock riproduttore nell'Atlantico occidentale è formato da individui più grandi che maturano più tardi.

Per quanto riguarda l'alimentazione *Thunnus thynnus* si nutre di una grande varietà di prede e per questo è stato definito predatore opportunistico. Le larve e i giovani più piccoli si cibano

principalmente di zooplancton, con una prevalenza di copepodi, mentre la dieta dei giovani più adulti e quella degli individui adulti è rappresentata principalmente da pesci, cefalopodi (soprattutto calamari) e crostacei. Questi gruppi includono numerose specie e la composizione è determinata principalmente dalla zona di caccia. Gli individui possono inseguire e consumare prede individuali, ma la strategia più utilizzata è quella del ram-feeding (nuotare in mezzo a banchi di prede a bocca aperta).

Sebbene le migrazioni riproduttive del tonno fossero note sin dall'antichità con le osservazioni condotte da Aristotele e altri filosofi del tempo e i ricercatori siano a conoscenza dei movimenti migratori trans-atlantici da decenni (Sella 1927, Sella 1929b), un aumento significativo delle conoscenze relative ai movimenti e alla struttura degli stock di tonno rosso è scaturito dall'utilizzo delle marcature elettroniche (rilevazione satellitare) applicate ai singoli individui. Studi recenti condotti con questo approccio hanno indicato che il mescolamento tra i due stock è maggiore rispetto a quanto assunto in precedenza. Inoltre questi nuovi approcci hanno fornito informazioni preziose sulla dinamica migratoria degli individui all'interno dell'area di distribuzione ed in particolare sui movimenti stagionali legati all'alimentazione e sui movimenti orizzontali e verticali nell'ambiente marino.

I dispositivi utilizzati nelle marcature elettroniche consentono di registrare diversi parametri ambientali come l'intensità luminosa, la temperatura dell'acqua e corporea e la pressione ad intervalli frequenti. Questi dati vengono poi utilizzati per ricostruire la posizione geografica in associazione alla profondità a cui nuota l'animale e alla fisiologia termica. Inoltre i dati sulla temperatura e la profondità forniscono informazioni sugli habitat preferiti, sulla dieta e i siti di nutrizione e sulla biologia riproduttiva. Block *et al.* (Block *et al.* 2001a, Block *et al.* 2005) hanno raccolto dati relativi ai movimenti del tonno rosso nel periodo tra il 1999 e il 2004 mediante marcature elettroniche. Grazie a queste informazioni hanno identificato pattern distinti della distribuzione del tonno rosso che sembra essere associata all'ontogenesi, ai cambiamenti stagionali e alla fedeltà ai siti di accoppiamento. Gli individui seguiti per meno di un anno, mostrano una notevole permanenza nell'Atlantico occidentale, muovendosi tra i siti invernali al largo degli stati meridionali delle coste nord americane (Nord Carolina) fino alle acque più a nord nel Golfo del Maine. Individui più piccoli di 200 cm si muovevano a nord nord-est ad inverno inoltrato o a inizio primavera, con la maggior parte dei pesci che si muovevano verso le acque del New England, occupando aree sovrapposte alla piattaforma continentale. Nessuno di questi pesci è penetrato nelle aree riproduttive. L'inverno seguente i pesci sono tornati nel nord Carolina. Viceversa i tonni più vecchi (>200 cm) hanno visitato i siti di riproduzione nel Golfo del Messico o nel Mediterraneo, sebbene nessun esemplare

abbia visitato entrambi i siti durante il periodo di osservazione. Per tre anni consecutivi cinque individui sono tornati nei siti di riproduzione, o nel Golfo del Messico o nel Mediterraneo. Infine, 23 individui marcati lungo le coste americane dell'Atlantico occidentale hanno effettuato migrazioni transatlantiche penetrando nel Mediterraneo. Tuttavia è da notare che gli individui riproduttori dell'Atlantico occidentale (quelli che visitano le aree di riproduzione del Golfo del Messico) non hanno mai raggiunto l'area di riproduzione dello stock orientale (cioè il Mediterraneo). Lo stesso pattern è mostrato dai riproduttori dello stock orientale (quelli che cioè visitano abitualmente il Mediterraneo nei mesi riproduttivi) in quanto non sono mai registrate migrazioni di questi individui all'interno del Golfo del Messico. L'analisi dei pattern di migrazione geografica del tonno rosso indicano univocamente che le aree di mescolamento sono lungo la costa orientale degli Stati Uniti e nelle aree di alimentazione del nord-centro Atlantico. Queste analisi hanno consentito di confermare la fedeltà ai siti di riproduzione: tonni giovani o subadulti marcati lungo le coste dell'Atlantico occidentale, una volta raggiunta la maturità sessuale, sono rientrati nel Mediterraneo e in seguito sono rimasti nelle zone di alimentazione dell'Atlantico orientale.

Anche nel Mediterraneo negli ultimi anni sono stati condotti progetti con l'utilizzo delle marcature elettroniche. Nel 2005 De Metrio *et al.* hanno pubblicato uno studio che riassume i risultati ottenuti dal rilascio di individui marcati nel periodo 1998-2004 in tre aree del Mediterraneo: il Mar Levantino nel Mediterraneo orientale, il mar Tirreno al largo di Corsica e Sardegna nel Mediterraneo centrale e dallo stretto di Gibilterra fino all'Atlantico orientale nel Mediterraneo occidentale. Nel periodo di osservazione degli individui non sono stati identificati movimenti transatlantici, anche se a volte sono necessari periodi più lunghi per identificarli. I movimenti degli esemplari registrati nel Mediterraneo sono spesso limitati, in particolar modo per gli individui marcati nelle aree centrale e orientale del bacino. Nel Mediterraneo orientale i tonni sono rimasti molto vicini al sito di rilascio, mentre quelli rilasciati nel Mediterraneo centrale ed occidentale presentano degli spostamenti più consistenti, con diversi esemplari che hanno oltrepassato lo stretto di Gibilterra (alcuni spingendosi anche migliaia di chilometri verso nord in prossimità dell'Islanda (Figura 4). De Metrio *et al.* (2005) ipotizzano che possa esistere un legame tra il pattern di movimento o la distanza degli spostamenti, e la taglia degli esemplari. In ogni caso, l'insieme dei dati suggerisce un'elevata complessità e diversificazione dei tratti ecologici e riproduttivi del tonno rosso.

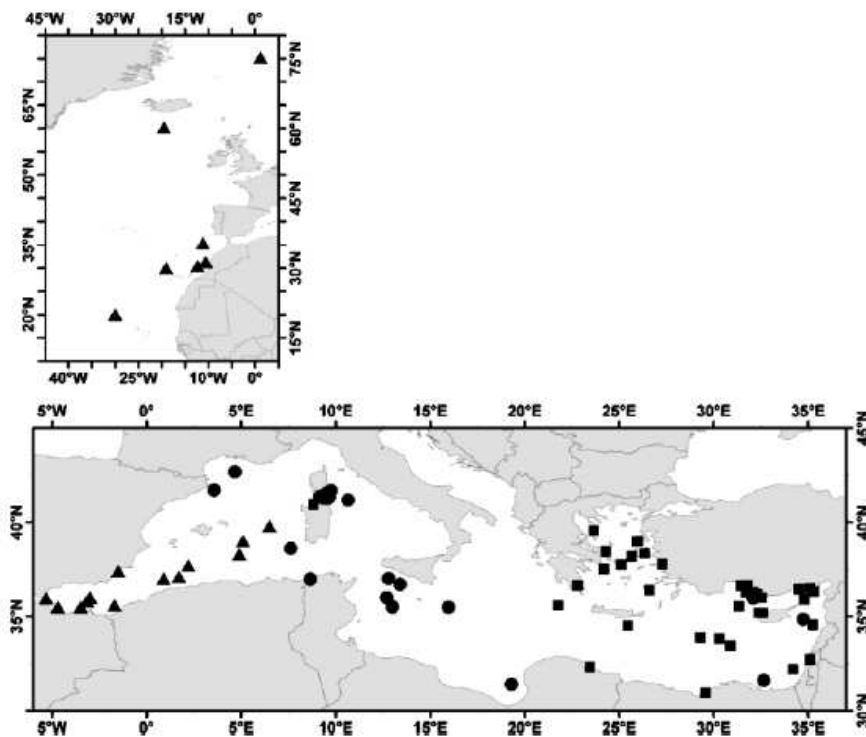


Figura 4. Posizioni delle tag elettroniche satellitari di ABFT distribuiti nel mar Mediterraneo e nell'est Atlantico durante 1998-2004. I diversi simboli rappresentano la distribuzione dei siti nel Mediterraneo: est (quadrato), centrale (cerchio), ovest Mediterraneo e est Atlantico vicino allo stretto di Gibilterra (triangolo). (De Metrio *et al.* 2005).

1.1.2. Analisi Genetiche

L'applicazione dell'analisi della variazione dei marcatori genetici per la risoluzione della struttura genetica di popolazione nel tonno rosso annovera un elevato numero di studi e ricerche, condotte con vari tipi di marcatori e loci e su vari campioni di tipo geografico. Dalla loro comparazione emergono pattern che spesso risultano contraddittori, anche in ragione dell'utilizzo di marcatori poco risolutivi (ma gli unici disponibili sulla base delle tecnologie del periodo) o di campionamenti limitati e senza un disegno sperimentale appropriato per la risoluzione del problema. In particolare, molti studi sono stati condotti per risolvere la separazione riproduttiva dei due stock dell'Atlantico orientale e occidentale. Assolutamente rari sono stati gli studi sulla variazione temporale delle popolazioni. Sulla base della scala temporale utilizzata questi studi possono dare indicazioni significative sulla stabilità o instabilità della struttura genetica in una data area geografica oppure sui cambiamenti

demografici delle popolazioni. La realizzazione di analisi temporali è tuttavia limitata dalla disponibilità di campioni storici idonei per l'analisi genetiche.

Inizialmente, come per molte altre specie di pesci marini, per la risoluzione della struttura genetica dei due stock sono stati applicati i loci allozimatici. Gli studi condotti con questo tipo di marcatori non sono riusciti però ad evidenziare alcuna suddivisione delle popolazioni di *T. thynnus* e, in generale, hanno mostrato un livello di diversità molto basso. Al contrario lo studio condotto usando la regione di controllo I (CR-I) del DNA mitocondriale (Alvarado Bremer *et al.* 1999), ha rivelato una sostanziale variabilità genetica ($h=0.99$), ed ha fornito la prima evidenza di suddivisione di popolazione tra la popolazione dell'Atlantico occidentale e quella del Mediterraneo ($\Phi_{ST}= 0.0119$; $P=0.027$). Queste analisi non sono state però riconfermate da un successivo studio condotto considerando un numero assai più ampio di campioni (Alvarado Bremer *et al.* 2005). La ragione di questa contraddizione dei risultati potrebbe risiedere nel fatto che i campioni rappresentativi dello stock occidentale di tonno rosso, collezionati al largo delle coste degli Stati Uniti, sono in realtà campioni formati da individui nati nelle due aree riproduttive in quanto, come dimostrato in modo risolutivo solo recentemente da Rooker *et al.* (2008) in queste aree avviene il mescolamento di individui dei due stock. In studi successivi (Carlsson *et al.* 2007) sono stati quindi utilizzati campioni differenti ed in particolare individui nati nell'anno (young-of-the-year; YOY) e adulti provenienti dal Golfo del Messico e dal Mediterraneo. Carlsson *et al.* (2007) hanno esaminato gli YOY campionati nelle aree di riproduzione del Golfo del Messico e del Mediterraneo, identificando un differenziamento statisticamente significativo con 8 loci microsatelliti ($F_{ST}= 0.0059$, $p<0.001$). Inoltre, anche il differenziamento genetico a carico della regione di controllo mitocondriale è risultato significativo ($\Phi_{ST}= 0.0129$; $P=0.014$). Questi risultati sono stati confermati (Boustany *et al.* 2006) anche mediante l'analisi di sequenze di DNA mitocondriale di esemplari adulti sia dal Golfo del Messico che dal Mediterraneo ($\Phi_{ST}= 0.011$; $P=0.041$). In generale i risultati di questi studi sostengono l'ipotesi di un isolamento genetico tra il *T. thynnus* del Golfo del Messico e del Mediterraneo e sono consistenti con il comportamento filopatrino osservato mediante gli studi condotti con marcature satellitari (Block *et al.* 2001, 2005).

Un solo studio significativo è stato condotto all'interno del Mediterraneo. Utilizzando tre campioni geografici replicati in anni successivi, Carlsson *et al.* (2004) hanno identificato una stabile e statisticamente significativa eterogeneità genetica fra campioni provenienti da tre regioni del Mediterraneo (Baleari, Tirreno meridionale e Ionio meridionale), utilizzando la variazione sia di sequenze della CR ($\Phi_{ST}= 0.0233$; $P=0.000$) che delle frequenze alleliche di

nove loci microsatelliti ($F_{ST}=0.0023$; $P=0.038$). Tuttavia, i risultati dell'analisi dei confronti a coppie di campioni risultano contraddittori al loro interno in quanto un differenziamento genetico statisticamente significativo è stato individuato solo tra il campione dello Ionio meridionale e quello del Tirreno. Tuttavia, sulla base di questi dati gli autori suggeriscono una strutturazione del tonno rosso nel Mediterraneo con l'esistenza di una popolazione discreta nel Mediterraneo orientale, della quale il campione dello Ionio potrebbe essere rappresentativo della parte più occidentale. L'evidenza di strutturazione nella popolazione del Mediterraneo complica ulteriormente l'ipotesi dell'esistenza dei due stock, mettendo in luce uno scenario ecologico e demografico molto complesso e ancora irrisolto, suggerendo la presenza di una popolazione caratterizzata anche dall'esistenza di sottopopolazioni locali. Evidenze recenti in tale senso sono state ottenute con le marcature elettroniche (De Metrio *et al.* 2005) in quanto hanno rivelato che all'interno del Mediterraneo (soprattutto nella parte orientale) i tonni attuano movimenti migratori limitati..

1.1.3. Sfruttamento del tonno rosso nel Mediterraneo nel XX° secolo

Dall'inizio del ventesimo secolo e in particolare dagli anni '80, la pesca del tonno rosso è aumentata in modo esponenziale. Da un lato il miglioramento delle tecniche di pesca grazie agli avanzamenti tecnologici, dall'altro lato lo sviluppo del mercato giapponese di sushi e sashimi hanno spinto lo sfruttamento della specie a livelli mai registrati prima.

La pressione da parte del mercato giapponese ha provocato un netto aumento dell'efficienza e della capacità delle industrie di pesca costituite durante gli anni '80 e '90: nel 1998 le navi tonniere con reti da circuizione erano due volte più lunghe e quattro volte più potenti rispetto al 1970. Inoltre erano equipaggiate con strumenti potenti per la ricerca dei banchi di tonno come radar, scandagli, sonar e aerei. La pesca del tonno rosso si è così espansa e attualmente viene svolta nell'intero bacino del Mediterraneo. L'aumento delle flotte è stato accompagnato da un rapido aumento delle catture dagli anni '80 e il Mediterraneo ha acquisito una posizione di dominanza come area di pesca del tonno rosso rispetto all'Atlantico. Negli anni '90 le catture nell'est Atlantico e nel Mediterraneo hanno raggiunto un picco storico superando le 50000 tonnellate (ICCAT 2007)(Figura 5).

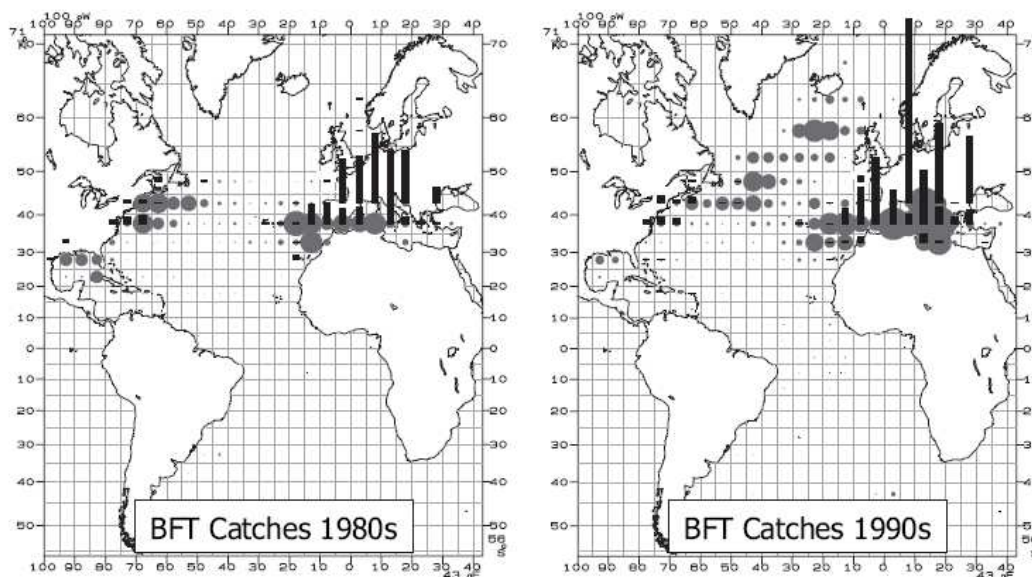


Figura 5. Catture totali di tonno rosso negli anni 1980-1989 e 1990-1999, per località geografica (quadrati 5x5) e per reti principali: le barre rappresentano le catture totali delle reti di superficie (sciabiche, barchini pasturatori, lenze a mano e trappole); i cerchi rappresentano le catture totali dei palamiti. Da Fromentin e Powers (2005).

La preoccupazione per lo stato di conservazione del tonno rosso è documentata già nei primi anni '70 quando l'ICCAT ha formulato le prime raccomandazioni per limitare lo sforzo di pesca e porre dei limiti sulla taglia massima di cattura consentita. Per molte flotte tonniere operanti nell'Atlantico orientale e Mediterraneo esistono enormi incertezze riguardo ai dati delle catture totali e della loro composizione in taglia. Tuttavia, la produttività dello stock orientale è stata stimata in circa 25000 tonnellate. Le associazioni internazionali per la conservazione delle biodiversità hanno cominciato ad interessarsi al tonno rosso a causa della drastica diminuzione di abbondanza registrata in modo particolare nello stock occidentale. Negli anni '90 è stato quindi sviluppato dall'ICCAT un metodo di regolazione del mercato del tonno per monitorare gli esemplari che venivano commercializzati nel mercato giapponese. Questo strumento di regolazione ha consentito di ridurre la quantità di catture non segnalate nello stock occidentale. Per quanto riguarda lo stock orientale, nel 2002 l'ICCAT ha stabilito la quota annuale di catture nel periodo 2002-2006 pari a 32000 tonnellate. Inoltre, il limite di taglia di 6.4 kg imposto nel 1975 è stato alzato a 10 kg per il Mediterraneo e 30 kg per l'ovest Atlantico (in base alla diversa età di maturazione). Tuttavia è stimato che le catture illegali, non regolamentate e non riportate di tonno rosso nello stock siano estremamente consistenti, arrivando ad eguagliare quelle regolari (Figura 6).

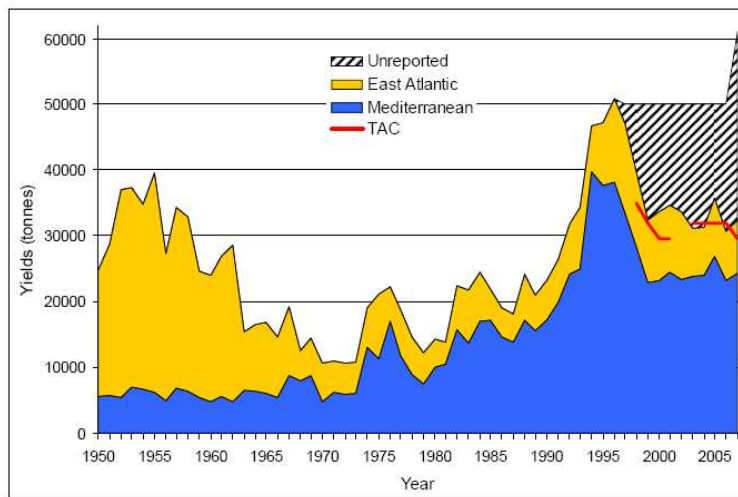


Figura 6. Catture riportate da dati della Task I dal 1950 al 2007 divisi per aree geografiche principali insieme alle catture non riportate stimate dal Comitato dalla capacità attiva di pesca e tasso medio di catture nell'ultimo decennio. (ICCAT 2007)

A peggiorare ulteriormente la situazione dello stock dell'est Atlantico e aumentare il numero di catture non segnalate, alla fine degli anni '90 sono comparse in Spagna le prime industrie di allevamento del tonno rosso (tuna farm). Questo tipo di pratica si è sviluppato rapidamente in tutto

il Mediterraneo negli ultimi venti anni con impianti anche in Italia, Libia, Turchia, Tunisia, Croazia, Cipro, Malta, Spagna e Francia (Figura 7). Questi impianti non consentono di contrastare il depauperamento della specie e migliorare la sua sostenibilità, in quanto i tonni che vengono allevati in enormi gabbie a mare aperto sono tonni prelevati dalle popolazioni naturali. Nelle gabbie gli individui vengono nutriti in modo da aumentare la quantità di grasso ed ottenere un prodotto più adatto al mercato giapponese. Lo scopo è perciò ingrassare gli esemplari. Quindi la produzione degli impianti di allevamento dipende totalmente dalle popolazioni naturali di tonno rosso e contribuiscono al loro impoverimento aumentando il numero di catture non segnalate e colpendo anche gli individui giovani che, in base ai limiti di taglia stabiliti, non potrebbero essere pescati. Inoltre queste gabbie possono essere spostate da un'area ad un'altra del Mediterraneo facendo perdere ogni tipo di controllo sulla tracciabilità degli individui allevati. Recentemente l'Unione Europea ha finanziato un programma di ricerca per la riproduzione in cattività del tonno rosso ma i primi tentativi sono falliti.

Le enormi incertezze riguardo i dati di pesca del tonno rosso nell'Atlantico orientale e nel Mediterraneo, insieme alla mancanza di conoscenze sui processi chiave delle dinamiche di popolazione (riproduzione, migrazione, struttura), hanno fortemente influenzato e a volte distorto la percezione del suo stato di conservazione. Tuttavia, nonostante le attuali incertezze, è opinione comunemente accettata che questa specie sia stata pesantemente sovrasfruttata per diversi decenni.

A)



B)



Figura 7. **A)** esemplari di tonno rosso (*Thunnus thynnus*). **B)** Gabbia in mare aperto utilizzata per l'allevamento del tonno rosso.

1.2. Cos'è la genetica della conservazione?

La genetica della conservazione consiste nell'utilizzo di teorie e tecniche genetiche per ridurre il rischio di estinzione delle specie minacciate. L'obiettivo a lungo termine è quello di preservare le specie quali entità dinamiche, capaci di affrontare i cambiamenti ambientali. La genetica della conservazione deriva dalla genetica evoluzionistica e dalle teorie della genetica quantitativa che sono alla base della riproduzione controllata di piante ed animali domestici. Queste teorie trattano generalmente di grandi popolazioni in cui la composizione genetica è governata da fattori deterministici (coefficienti di selezione, ecc.). La genetica della conservazione costituisce oggi una disciplina distinta che si concentra sulle conseguenze risultanti dalla riduzione di popolazioni una volta numerose e con esoincrocio, ad unità ridotte dove i fattori stocastici e gli effetti dell'inincrocio hanno un effetto preponderante.

Il campo della genetica della conservazione comprende anche l'uso di analisi genetiche molecolari per chiarire aspetti della biologia delle specie rilevanti per la gestione della conservazione.

I campi di applicazione di queste tecniche sono molteplici:

- riduzione del rischio di estinzione minimizzando l'inincrocio e la perdita di diversità genetica
- comprensione della struttura nelle popolazioni frammentate
- risoluzione delle ambiguità tassonomiche
- identificazione di specie o popolazioni a rischio a causa della ridotta variabilità genetica
- definizione delle unità di gestione all'interno della specie
- individuazione delle specie ibride
- scelta delle popolazioni migliori per la reintroduzione
- aspetti forensi
- migliore comprensione della biologia della specie.

In generale l'utilizzo delle analisi genetiche nel campo della conservazione cade in due aree principali: osservazione delle conseguenze genetiche di cali demografici sulla fitness media e la probabilità di estinzione; uso di dati derivanti da analisi genetiche per inferire aspetti della storia demografica delle popolazioni.

La diversità genetica è necessaria alle popolazioni per adattarsi ai cambiamenti ambientali. Le popolazioni di grandi dimensioni, le cui specie praticano naturalmente l'esoincrocio, possiedono una notevole quota di diversità genetica, ma questa è notevolmente ridotta nelle specie e nelle popolazioni in pericolo. Le popolazioni di interesse per la conservazione sono piccole e/o stanno diminuendo la loro numerosità. Le popolazioni piccole e isolate subiscono un'accelerazione dell'inincrocio e della perdita di diversità genetica dovuta alla deriva, cioè al fatto che nell'arco delle generazioni le frequenze alleliche fluttuano casualmente da una generazione a quella successiva e l'ampiezza delle fluttuazioni è funzione inversa del numero di individui che si riproducono. Questi fenomeni determinano una diminuzione della fitness riproduttiva (depressione da inincrocio) ed una ridotta capacità di evolvere in risposta a cambiamenti ambientali legata al depauperamento del pool allelico. La gestione genetica *in situ* delle specie in pericolo comporta il recupero di queste piccole popolazioni, la gestione di popolazioni frammentate, l'attenuazione dell'impoverimento genetico dovuto all'ibridazione con specie affini e la minimizzazione degli impatti deleteri del prelievo. L'analisi di vitalità della popolazione può essere utilizzata sia per quantificare il livello di minaccia che per confrontare opzioni alternative di intervento.

1.2.1. Diagnosticare i problemi genetici

Una necessità preliminare per la gestione della diversità genetica delle popolazioni naturali consiste nella diagnosi del loro status. È necessario determinare quanto è grande la popolazione e quanti individui sono fertili (N_e), se ha subito nel passato eventi significativi di colli di bottiglia, l'avvenuta perdita di diversità genetica, la presenza di segnali di inincrocio e di struttura di popolazione.

La necessità di preservare la diversità genetica è una delle priorità della conservazione poiché è collegata a due aspetti fondamentali nella gestione delle specie. Primo, il cambiamento ambientale è un processo continuo e la diversità genetica è necessaria alle popolazioni per evolvere ed adattarsi a tale cambiamento. Secondo, la perdita di diversità genetica è di solito associata all'inincrocio e alla riduzione complessiva della riproduzione e sopravvivenza (fitness). Le popolazioni che hanno subito un collo di bottiglia mostrano tipicamente maggiori livelli di inincrocio, diminuita fitness riproduttiva, ridotta diversità genetica e compromessa capacità di evoluzione. Il problema delle piccole popolazioni riguarda la perdita di alleli, soprattutto quelli rari, da una generazione a quella successiva che potrebbero non essere trasmessi per puro effetto del caso.

L'azione della deriva genetica sulle popolazioni di piccole dimensioni determina due effetti principali: i) perdita di diversità genetica e fissazione degli alleli all'interno delle popolazioni, con conseguente riduzione del potenziale evolutivo; ii) diversificazione tra popolazioni che derivano tutte dalla stessa popolazione di origine (sviluppo di struttura).

In popolazioni chiuse, cioè che non sono sottoposte a immigrazione, l'incrocio tra individui imparentati (inincrocio) è inevitabile. Con il tempo, se non intervengono altri fenomeni, tutti gli individui diventeranno imparentati e gli incroci tra individui non correlati saranno impossibili. Questo è la conseguenza del piccolo numero di fondatori e della ridotta dimensione della popolazione. L'inincrocio è di estrema importanza per la biologia della conservazione poiché porta a riduzioni dell'eterozigosi, a diminuzione delle potenzialità riproduttive e della sopravvivenza (fitness riproduttiva) e ad un aumento del rischio di estinzione. La perdita di fitness riproduttiva come conseguenza dell'inincrocio è definita depressione da inincrocio.

Per stimare il livello di inincrocio di una popolazione si utilizza il coefficiente di inincrocio (F) che rappresenta la probabilità che in un individuo gli alleli per un certo locus siano identici per discesa. L'inincrocio aumenta i livelli di omozigosi, non altera le frequenze alleliche ma cambia quelle genotipiche. Una delle principali conseguenze pratiche è che gli omozigoti per gli alleli recessivi deleteri diventano più frequenti rispetto alle popolazioni con accoppiamento casuale. Questa è la causa principale della depressione da inincrocio. Nelle specie che praticano esoincrocio, l'inincrocio si verifica prevalentemente quando le popolazioni sono di dimensioni ridotte.

Tutte le conseguenze genetiche avverse che si verificano in una popolazione di piccole dimensioni dipendono dalla dimensione effettiva (N_e) della popolazione piuttosto che dal numero assoluto degli individui, ossia la sua dimensione assoluta o censita (N_c). Tra gli individui in grado di riprodursi esiste poi una considerevole variazione del contributo alle generazioni successive, che si traduce anch'esso in fluttuazioni non adattative delle frequenze alleliche. Inoltre altri fattori possono contribuire ad aumentare il divario tra il numero di individui di cui è composta una popolazione e quelli che determinano la costituzione genetica della generazione successiva. E' quindi importante definire cosa si intende con dimensione della popolazione in genetica della conservazione. Di solito le popolazioni reali vengono confrontate con una popolazione ideale di cui si conoscono bene le caratteristiche in quanto definite da assunzioni che ne semplificano le dinamiche e che hanno consentito di derivare gran parte della teoria riguardante lo studio della genetica di popolazione. Di conseguenza è

possibile confrontare le popolazioni reali con quelle ideali usando la teoria per ottenere delle previsioni pratiche. Questa popolazione ideale possiede dimensione costante, generazioni non sovrapposte, assenza di migrazione o flusso genico, riproduzione casuale, nessuna selezione e mutazione trascurabile. Per dimensione effettiva di una popolazione si intende perciò la dimensione di una popolazione ideale che va incontro ad una perdita di diversità genetica della stessa intensità di quella che si verifica nella popolazione reale. Quindi il valore di N_e di una popolazione è la misura del suo comportamento genetico relativo a quello di una popolazione ideale.

Qualsiasi caratteristica di una popolazione reale che si scosta da quelle di un popolazione ideale determinerà una differenza tra N_e e N_c . In generale le popolazioni reali non seguono le assunzioni definite per quelle ideali per cui spesso si osservano delle deviazioni che portano a $N_e < N_c$. La N_e può essere stimata da dati demografici sul rapporto tra i sessi, varianza della dimensione della famiglia e fluttuazioni della dimensione della popolazione nel corso delle generazioni, oppure da dati genetici.

Cospicue riduzioni del rapporto N_e/N_c derivano dalle fluttuazioni della dimensione della popolazione nel corso delle generazioni. Le popolazioni naturali variano di numerosità in conseguenza di variazioni che si verificano nell'ambiente in cui vivono, come cambiamenti nella disponibilità di risorse alimentari, nelle condizioni climatiche, le epidemie, le catastrofi, le predazioni...La dimensione effettiva di una popolazione che fluttua non corrisponde alla media aritmetica, ma alla media armonica di N_e nel corso delle generazioni ed il valore calcolato con la media armonica è più vicino al valore minimo di N_e registrato nell'intervallo di generazioni considerato. Questa relazione è importante dal punto di vista della conservazione delle specie in quanto un allele perso in una generazione da una popolazione di dimensioni ridotte non viene riguadagnato quando la popolazione si ristabilisce e lo stesso avviene per l'inincrocio che non si riduce immediatamente quando la popolazione aumenta di dimensione.

Le sequenze del DNA contengono informazioni sulla storia evolutiva di una specie, la dimensione delle popolazioni, la loro eventuale frammentazione, le diverse forme di selezione naturale...L'analisi delle differenze nella sequenza del DNA tra individui e popolazioni permette di esplorare i processi evolutivi e gli eventi demografici di una specie avvenuti nel passato. Metodi basati sulla coalescenza e le genealogie dei geni permettono di ottenere queste informazioni.

1.3. Alberi genealogici, teoria della coalescenza e analisi dei polimorfismi genetici (Rosenberg and Nordborg 2002)

La teoria della coalescenza è basata sul concetto che partendo dalle attuali sequenze di DNA ad un determinato locus in una popolazione, sia possibile risalire indietro nel tempo fino al loro antenato comune. Altri alleli presenti nel passato possono essere stati persi per deriva genetica o selezione, mentre nuovi alleli si sono generati attraverso mutazione lungo la genealogia del gene. Il pattern evolutivo dell'attuale distribuzione di alleli per un locus può essere rappresentato come rami di un albero che convergono indietro nel tempo fino ad una singola sequenza allelica ancestrale.

I pattern di coalescenza vengono di solito rappresentati come genealogie del gene, che mostrano andando indietro nel tempo l'albero delle linee di discendenza presenti nella popolazione. Ogni linea di discendenza nel campione attuale consiste in un tratto di DNA ad uno specifico locus. La posizione dei nodi (eventi di coalescenza) e la lunghezza dei rami dell'albero rappresentano le origini e gli intervalli di tempo che hanno determinato i pattern di diversità genetica osservati. La genealogia del gene è sconosciuta, ma, in assenza di selezione, gli intervalli tra eventi di coalescenza seguono leggi probabilistiche note (Kingman 1982a) e dipendono in sostanza da due fattori: dalle dimensioni del campione (maggiori le dimensioni, maggiore la probabilità di un evento) e da quella della popolazione (maggiori queste dimensioni, minore la probabilità che due sequenze abbiano lo stesso antenato). Tramite questi concetti è possibile generare per simulazione un grande numero di possibili genealogie, da cui ricostruire la storia evolutiva dei tratti di DNA campionati e quindi anche dei diversi alleli, allo stesso modo in cui si rintraccia l'origine, o la perdita, di alleli utilizzando i pedigree.

Considerando quindi un locus nel genoma di una specie, tutti gli alleli sono correlati gli uni agli altri e ad un antenato comune più recente (MRCA) attraverso una singola genealogia. I polimorfismi presenti in questo locus sono dovuti a mutazioni avvenute lungo i rami di questa genealogia e la frequenza di ogni variante di sequenza è determinata dalla frazione di rami che ereditano la variante. Il pattern dei polimorfismi riflette perciò sia la storia della coalescenza delle linee genealogiche, che danno origine all'albero, sia la storia mutazionale.

La struttura delle genealogie dipende da diversi fattori, per cui anche riuscendo a ricostruire una genealogia corretta, sono diversi i processi demografici che possono dare origine alla stessa genealogia e, d'altra parte, genealogie diverse possono essere ricondotte ad uno stesso processo. Quindi la ricostruzione delle genealogie può essere trattata come un processo

stocastico ed è necessario incorporare queste fonti di variabilità nell'analisi dei dati. E' possibile formulare delle assunzioni relative al processo che ha dato origine ai dati e immaginare molte ripetizioni casuali del processo evolutivo. Se la frazione di storie casuali genealogiche e mutazionali che dovrebbero dare origine ai dati osservati è piccola, possiamo concludere che le assunzioni fatte nel simulare i diversi modelli genealogici non riescono a spiegare il nostro campione attuale. Per stimare le genealogie che possono essere generate dai diversi processi evolutivi, sono necessari modelli che consentano di costruire delle genealogie casuali: il coalescente è uno di questi modelli.

Rendendo più complesse le assunzioni, tutte le altre forze evolutive (ricombinazione, selezione, selezione naturale, migrazione) possono essere incorporate nella struttura genealogica. L'effetto principale della ricombinazione è il fatto che siti del genoma in linkage possiedono una struttura genealogica diversa. La ricombinazione è un fattore molto importante nell'inferenza evolutiva, perché loci non legati, o non strettamente legati, possono essere considerati come repliche indipendenti dei processi evolutivi. In assenza di ricombinazione, il genoma intero potrebbe essere descritto con una unica genealogia, e non si avrebbe mai più di una singola replica indipendente. Perciò incorporare la ricombinazione tra loci all'interno delle strutture genealogiche consente di considerare ogni locus indipendente dagli altri (è un'assunzione non completamente vera, ma rappresenta una buona approssimazione) e permette una modellizzazione più efficiente, dal punto di vista statistico, del processo evolutivo. La precisione dei metodi di inferenza evolutiva aumenta rapidamente con il numero di geni studiati (loci) e lentamente con il numero degli individui campionati proprio perché si ottiene informazione indipendente da genealogie differenti.

L'idea alla base della teoria della coalescenza è che in assenza di selezione, le linee genealogiche attuali estraggono a caso i loro genitori, immaginando di spostarsi indietro nel tempo. Ogni volta che due linee estraggono gli stessi genitori coalescono e questo processo continua fino a che tutte le linee si uniscono in una singola linea che rappresenta l'antenato comune più recente. Come detto in precedenza, il tasso a cui le linee coalescono dipende dalla dimensione del campione (più linee, tasso più veloce), e dalla grandezza della popolazione (più genitori tra cui scegliere, più lento il tasso; Figura 8).

Molti altri fattori possono essere inclusi nel modello. Alcuni come la variazione nel successo riproduttivo, suddivisione della popolazione per classi di età e rapporto sbilanciato dei sessi cambiano solo il tasso di coalescenza, altri come la struttura di popolazione e fluttuazioni nella dimensione della popolazione cambiano anche la forma delle genealogie.

E' fondamentale fare una distinzione tra metodi genealogici e metodi filogenetici. Entrambi prevedono l'uso di alberi ma sono molto diversi. I metodi filogenetici stimano alberi, sono stati sviluppati per determinare il pattern di discendenza delle specie, che si assume abbia una forma simile ad un albero. Gli alberi stimati vengono usati per trarre conclusioni relative alle relazioni tra specie o tra le loro popolazioni o individui. I metodi genealogici non stimano gli alberi, li utilizzano per stimare i parametri relativi ai processi genealogici casuali che hanno dato origine ad ogni albero. L'approccio genealogico fornisce quindi una struttura statistica coerente in cui considerare ricombinazione, migrazione, selezione e altri processi.

1.3.1. Modello standard di coalescenza

Il modello più semplice di coalescenza è stato descritto per la prima volta da Kingman (Kingman 1982a, Kingman 1982b, Kingman 1982c): in questo modello le linee ancestrali, che rappresentano una serie di antenati genetici dei campioni attuali ad un locus, vengono disegnate immaginando di muoversi indietro nel tempo. La storia di un campione di dimensione n comprende $n - 1$ eventi di coalescenza, perché ogni evento diminuisce il numero di linee originarie di 1. Questo processo è quindi descritto da un numero n di linee nel presente che in seguito ad una serie di passaggi diminuiranno da n a $n - 1$, da $n - 1$ a $n - 2$...fino al passaggio finale che vede due linee coalescere in una unica, ossia all'antenato comune più recente (MRCA) dell'intero campione (Figura 9).

Figura 8

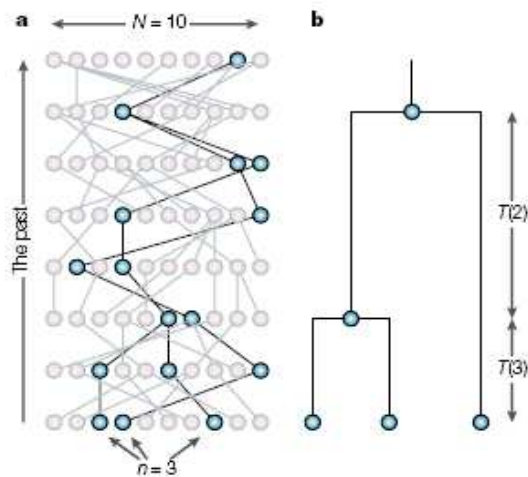


Figura 9

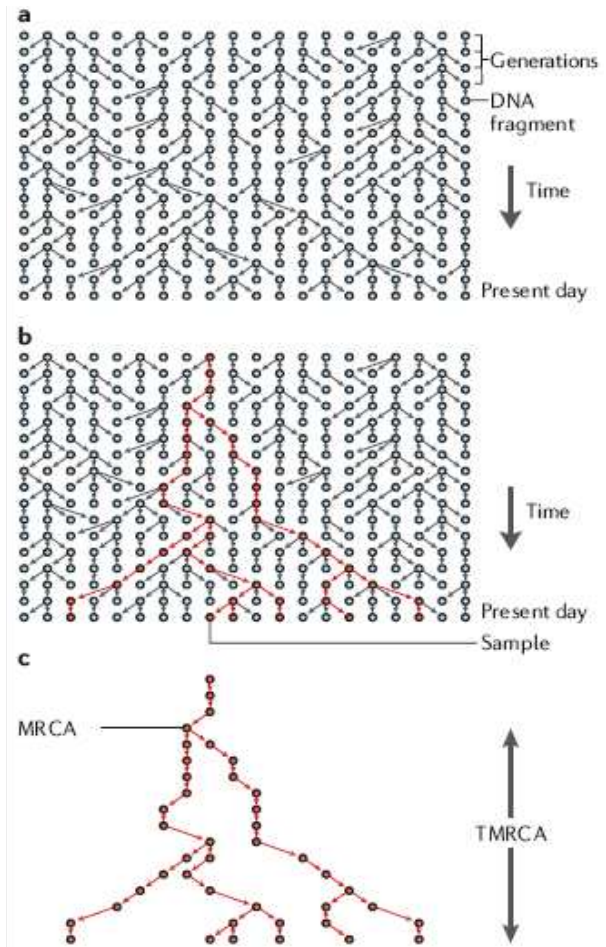


Figura 8. Principio base della teoria della coalescenza. a) Genealogia completa per una popolazione di 10 individui diploidi (popolazioni diploidi di N individui vengono generalmente studiate usando un modello aploide con $2N$ individui). Le linee nere tracciano la discendenza di tre linee campionate indietro nel tempo fino all'antenato comune. b) Sottogenealogia delle tre linee campionate. Nella versione base del coalescente è necessario registrare i tempi tra eventi di coalescenza ($T(3)$ e $T(2)$) e la sua topologia. N , numero di copie alleliche nella popolazione; n , dimensione del campione. (Da Rosenberg and Nordborg 2002)

Figura 9. Il coalescente senza ricombinazione. Le figure a-c illustrano l'idea alla base della teoria della coalescenza usando una popolazione di frammenti di DNA che evolvono secondo il modello Wright-Fisher, quindi in assenza di ricombinazione, in una popolazione di dimensione costante. Nel pannello a è mostrata l'evoluzione di una popolazione. In questa rappresentazione semplificata ogni riga corrisponde ad una singola generazione e ogni cerchio azzurro rappresenta un frammento in quella generazione. Le generazioni vengono rimpiazzate interamente dai discendenti, con le frecce che si muovono dai frammenti parentali a quelli dei discendenti. Il tempo attuale è rappresentato dall'ultima riga, ogni riga precedente rappresenta una generazione indietro nel tempo. Nel pannello b è indicata la genealogia di un campione attuale. In questo esempio sei frammenti, indicati in rosso, sono campionati dalla generazione corrente. La genealogia di questo campione viene quindi tracciata indietro nel tempo ed è indicata in rosso. Il pannello c sottolinea una delle caratteristiche chiave del coalescente: tutta l'informazione al di fuori della genealogia del campione di interesse può essere ignorata. Il coalescente fornisce una descrizione matematica della discendenza di un campione. Come ci spostiamo indietro nel tempo il numero di linee genealogiche diminuisce fino ad averne una unica. Questa linea rappresenta l'antenato comune più recente (MRCA), mentre il tempo in cui appare l'MRCA è definito tempo dall'antenato comune più recente (TMRCA) (Marjoram and Tavaré 2006).

Ad ogni evento di coalescenza due linee si fondono in una linea ancestrale comune, alla fine si ottiene un albero con ramificazioni in cui il tempo T_i rappresenta il momento in cui sono presenti esattamente i linee ancestrali rispetto al campione attuale. Il coalescente è un processo stocastico che include una struttura ad albero discreta e $n - 1$ intervalli temporali di coalescenza. Lo spazio delle genealogie è rappresentato dall'insieme di tutti i possibili alberi con i nodi e i nodi terminali ordinati nel tempo e con tempi di coalescenza $0 < T_i < \infty$ per un numero di linee $2 \leq i \leq n$.

Qualsiasi particolare genealogia o realizzazione del processo di coalescenza, specificherà il tipo di ramificazioni che descrivono le relazioni tra i membri all'interno del campione e i tempi di coalescenza. Le genealogie devono essere trattate usando regole statistiche, anche se rispetto agli eventi statistici, non possono essere osservate direttamente. Le informazioni relative alla storia genealogica vengono inferite dalle configurazioni dei polimorfismi in un campione, che a loro volta sono il risultato di un altro processo casuale: la mutazione.

Nei casi più semplici i modelli che consentono di descrivere la distribuzione di probabilità sia degli alberi genealogici che dei tempi di coalescenza assumono che le differenze genetiche non hanno conseguenze sulla fitness, per cui non ci saranno individui che presentano un vantaggio selettivo rispetto ad altri, queste differenze sono considerate neutrali e quindi non sono influenzate dalla selezione naturale.

Il numero di discendenti tra gli individui di una popolazione è una variabile casuale scambiabile, ovvero i discendenti sono distribuiti in modo identico, ma non necessariamente indipendente in quanto il loro numero totale è fissato (la grandezza della popolazione è costante).

Kingman (1982a,b) ha mostrato nei suoi studi che quando la dimensione della popolazione tende ad infinito, i tempi di coalescenza T_i sono indipendenti e distribuiti esponenzialmente come

$$f_{T_i}(t_i) = \binom{i}{2} e^{-\binom{i}{2} t_i} \quad t_i \geq 0, \quad i=2, \dots, n$$

In un campione grande, molti eventi di coalescenza si verificheranno in un periodo di tempo molto breve nella storia recente del campione e la conformazione di circa metà genealogia è determinata dalla coalescenza delle ultime due linee. Poiché i tempi di coalescenza sono mutualmente indipendenti, si ha

$$f_{T_n, \dots, T_2}(t_n, \dots, t_2) = \prod_{i=2}^n f_{T_i}(t_i)$$

Inoltre ogni coppia di linee ha la stessa probabilità di coalescere ad ogni evento di coalescenza. Questo significa che ogni struttura genealogica è ugualmente probabile: dal modello di coalescenza scaturiscono due proprietà importanti, ovvero la natura casuale delle unioni e delle ramificazioni degli alberi ed i tempi di coalescenza distribuiti in modo esponenziale come eventi indipendenti.

La distribuzione esponenziale dei tempi di coalescenza può essere descritta da un processo di Poisson in cui ognuna delle $i(i-1)/2$ coppie possibili coalescono indipendentemente con tasso $\lambda=1$.

1.4. La statistica Bayesiana in genetica (Beaumont and Rannala 2004)

La genetica ha sviluppato sofisticate teorie matematiche per affrontare i complessi processi evolutivi e demografici che portano alla diversità genetica. Usando questi strumenti teorici è spesso possibile costruire modelli realistici che riescono a spiegare i dati alla luce dei processi coinvolti. Riuscire a formulare questi modelli è il primo passo per lo studio dei processi basilari e fornisce i presupposti per le inferenze statistiche. La maggior parte delle proprietà genetiche di individui, popolazioni o specie sono il prodotto di forze di tipo stocastico che non possono essere studiate senza l'uso di modelli probabilistici. Questi processi così complessi che danno origine ai dati genetici sono il risultato di molti meccanismi che possono produrre gli stessi dati osservati ed è quindi fondamentale stabilire il modo migliore per scegliere quale, tra le possibili cause, abbia prodotto i dati osservati.

L'inferenza Bayesiana permette di affrontare problemi di questo tipo, ossia modelli caratterizzati da molti parametri interdipendenti. Una delle caratteristiche allettanti dell'analisi Bayesiana è la possibilità di incorporare informazioni (informazioni *a priori*) di cui si dispone nel momento in cui si deve specificare un modello.

1.4.1. Principi di inferenza Bayesiana

Secondo la statistica Bayesiana non esiste una distinzione logica tra i parametri del modello e i dati. Entrambi sono variabili casuali descritte da una distribuzione di probabilità congiunta specificata da un modello probabilistico. I ‘dati’ rappresentano le variabili osservate mentre i ‘parametri’ le variabili non osservate. La distribuzione congiunta è semplicemente il prodotto della likelihood e delle probabilità *a priori* dei parametri (*prior probability*). La *prior* incorpora l’informazione relativa ai valori dei parametri disponibile prima di esaminare i dati attraverso una distribuzione di probabilità. La likelihood è una distribuzione condizionale che specifica la verosimiglianza dei dati osservati dato un qualsiasi valore dei parametri ed è basata su un modello che descrive il processo. Queste due funzioni insieme combinano tutte le informazioni disponibili sui parametri. La statistica Bayesiana implica semplicemente la manipolazione di questa distribuzione congiunta allo scopo di produrre inferenze sui parametri, o sul modello di probabilità dati i valori osservati. Perciò lo scopo principale della inferenza Bayesiana è calcolare la distribuzione a posteriori dei parametri, ossia la distribuzione dei parametri condizionata dai dati osservati.

I metodi Bayesiani utilizzano quindi la likelihood per effettuare inferenze sui parametri del modello, ma consentono l’incorporazione di una ‘prior information’ relativa ai valori dei parametri. Dal punto di vista più formale, questa informazione deve essere espressa specificando una distribuzione a priori o pre-dati, $P(\theta)$, relativa ai parametri (θ). La distribuzione $P(\theta)$ è pesata verso quei valori di θ che sono considerati più probabili in accordo con l’informazione a priori. In assenza di qualunque informazione, le priors hanno distribuzione rettangolare nell’intervallo tra i valori di minimo e massimo del parametro (flat priors). Questa informazione a priori viene combinata con la likelihood, moltiplicandole insieme per ottenere la distribuzione a posteriori o post-dati $P(\theta/D)$:

$$P(\theta / D) = L(\theta)P(\theta) / P(D)$$

Da questa equazione si capisce facilmente che $P(\theta/D)$ sarà elevata per quei valori di θ che sono ben supportati dai dati (hanno una likelihood elevata) e sono coerenti con l’informazione a priori ($P(\theta)$ elevata). La distribuzione a posteriori rappresenta la confidenza relativa ai parametri stimati, tenendo conto sia dell’informazione a priori sia dei dati osservati.

La stima puntuale di un parametro si ottiene considerando alcune proprietà della distribuzione a posteriori, di solito la moda o la mediana della distribuzione. L’intervallo di stima di un parametro si ottiene considerando un insieme ‘credibile’ di valori (un insieme o intervallo che

contiene il vero parametro con probabilità $1-\alpha$, dove α rappresenta il livello di significatività, di solito 0.05).

Un approccio alternativo non-Bayesiano, alla statistica inferenziale include il metodo della *maximum likelihood* ed il metodo dei momenti, che formano le basi dell'inferenza classica o frequentista. Il metodo di maximum likelihood basa le inferenze interamente sulla funzione di likelihood, senza incorporare nessuna informazione a priori e scegliendo stime puntuali dei parametri che massimizzano la probabilità dei valori osservati dati i parametri del modello.

Esistono molte ragioni pratiche per usare le inferenze Bayesiane: se ad esempio il modello di probabilità che stiamo utilizzando include molte variabili interdipendenti limitate in un intervallo particolare di valori, il metodo della maximum likelihood richiede una massimizzazione multidimensionale, limitata al range di valori, allo scopo di individuare un insieme di parametri combinati che possano massimizzare la funzione di likelihood. In generale questo rappresenta un problema di analisi numerica difficile e potrebbe richiedere uno sforzo computazionale enorme. Nell'inferenza Bayesiana la prior impone automaticamente dei limiti ai valori che i parametri possono assumere e le inferenze relative ai valori dei parametri sulla base della distribuzione a posteriori di solito richiedono l'integrazione, piuttosto che la massimizzazione. Inoltre, metodi numerici sviluppati negli anni '50 che prevedono l'utilizzo dei metodi Markov Chain Monte Carlo (MCMC), applicabili su computer più potenti, hanno facilitato moltissimo la stima delle probabilità a posteriori consentendo anche calcoli per modelli genetici complicati, impossibili con gli approcci classici o con la maximum likelihood.

1.4.2. Metodi Monte Carlo Markov Chain

Con Monte Carlo Markov Chain si definisce una classe di metodi che consistono nel simulare un tipo particolare di processi stocastici, conosciuti come Catene di Markov, per studiare distribuzioni di probabilità complicate che non possono essere studiate in modo semplice utilizzando metodi analitici (Ross 1997). Una catena di Markov genera una serie di variabili casuali tali che la distribuzione di probabilità degli stati futuri è determinata completamente dallo stato attuale in qualsiasi punto della catena. Sotto determinate condizioni, una catena di Markov avrà una distribuzione stazionaria, perciò se la catena ha compiuto un numero sufficiente di iterazioni, gli stati che visita tenderanno ad una distribuzione di probabilità specifica che non dipende più dal numero di iterazioni o dallo stato iniziale della variabile. L'idea che sta alla base di tutti i metodi MCMC è costruire una catena di Markov con una

distribuzione stazionaria che rappresenta la distribuzione di probabilità di interesse, e campionare da questa distribuzione per trarne inferenze relative ai parametri. Nell'analisi Bayesiana questa distribuzione di solito è la distribuzione a posteriori congiunta di uno o più parametri. La forma più semplice di MCMC è l'integrazione Monte Carlo.

1.4.3. Integrazione Monte Carlo

L'idea che sta alla base dell'integrazione Monte Carlo (MC) è che le proprietà delle variabili casuali possono essere studiate simulando molte variabili e analizzando i risultati (Ripley 1987). Ogni replica delle simulazioni MC è indipendente, quindi la procedura è equivalente ad estrarre campioni ripetuti da una catena di Markov stazionaria in punti sufficientemente separati da non risultare correlati. L'integrazione MC è stata ampiamente applicata alla genetica statistica (Hudson 1990). Il metodo di simulazione MC ha il vantaggio che le stime ottenute non sono distorte e l'errore standard può essere stimato in modo accurato perché le variabili casuali simulate sono indipendenti e identicamente distribuite. Uno svantaggio è rappresentato dal fatto che con variabili multidimensionali complesse, che hanno uno spazio degli stati ampio (ad esempio un ampio intervallo di valori possibili), può essere necessario un numero di simulazioni enorme per ottenere stime accurate dei parametri.

1.4.4. Algoritmo Metropolis-Hastings

L'algoritmo Metropolis-Hastings (MH) (Metropolis et al. 1953, Hastings 1970) è simile alla procedura di simulazione MC poiché mira a campionare da una catena di Markov stazionaria per simulare osservazioni da una distribuzione di probabilità. Questo metodo tuttavia invece di simulare osservazioni indipendenti dalla distribuzione stazionaria, simula valori sequenziali dalla catena fino alla convergenza, quindi campiona valori simulati a determinati intervalli dalla catena per imitare campionamenti indipendenti dalla distribuzione stazionaria. L'algoritmo MH ha il vantaggio che può migliorare l'efficienza delle simulazioni quando lo spazio degli stati è ampio perché concentra le variabili simulate su valori con elevata probabilità nella catena stazionaria. Gli svantaggi includono il fatto che nella maggior parte delle applicazioni pratiche, non sono disponibili metodi rigorosi per determinare quando la catena ha raggiunto la convergenza o quali sono gli intervalli ottimali tra i campioni per estrarre la maggiore informazione possibile preservando al medesimo tempo l'indipendenza tra le osservazioni. L'algoritmo MH aumenta perciò l'efficienza della ricerca, al prezzo dell'introduzione di un maggior tasso di arbitrarietà.

1.4.5. Applicazioni alla genetica di popolazioni

Uno degli scopi della genetica di popolazioni è comprendere i meccanismi che hanno dato origine al pattern di variabilità molecolare osservata. Grazie ai recenti avanzamenti nel campo delle tecnologie di sequenziamento e genotipizzazione, è stato possibile ottenere notevoli quantità di dati che descrivono la variabilità genetica in diversi organismi. Dal punto di vista metodologico l'interesse principale di questo campo è stato lo sviluppo dettagliato di modelli stocastici allo scopo di descrivere l'evoluzione delle frequenze alleliche ad un locus nel tempo. Per modello si intende una formulazione matematica relativamente semplice dei processi biologici che hanno prodotto i dati e incorpora i parametri di interesse, come la mutazione o il tasso di ricombinazione. Tutti i modelli formulati per descrivere i processi biologici sono di natura stocastica, cioè comprendono l'effetto di fattori aleatori, non deterministici.

In origine, in genetica di popolazioni i modelli hanno permesso di predire come i pattern di variabilità sarebbero influenzati da forze quali deriva genetica, selezione, migrazione e ricombinazione. La teoria della coalescenza consente di formalizzare in termini probabilistici questi modelli che di solito sono il risultato di simulazioni effettuate considerando scenari variabili (scenari in cui variano i parametri del modello). Questo tipo di simulazioni consentono di esplorare l'effetto che ogni cambiamento dei parametri ha sui dati osservati.

Il coalescente può essere usato in diversi contesti e il modo più semplice è utilizzarlo come strumento di simulazione. Fornisce infatti un metodo efficiente e conciso per simulare insiemi di dati multipli sotto diversi scenari evolutivi plausibili. A questo punto, ottenuti i dati per simulazione, uno dei classici problemi della statistica classica è rappresentato dalla stima di parametri quali mutazione, migrazione, ricombinazione e tasso di crescita. A questo scopo è stato spesso usato il metodo della *maximum likelihood* che come già sottolineato precedentemente non rappresenta la scelta migliore nel momento in cui è necessario calcolare le probabilità congiunte di un insieme ampio di parametri (la distribuzione di probabilità congiunta è la distribuzione di probabilità di tutte le combinazioni di due o più variabili casuali). La migliore alternativa al problema di identificare la combinazione di parametri che massimizza la funzione di likelihood, il che richiede uno sforzo computazionale enorme, è utilizzare i metodi Bayesiani che misurano il volume sotto la superficie della probabilità a posteriori. In pratica con i metodi Bayesiani i parametri di interesse vengono stimati marginalizzando i parametri di disturbo e quello che si ottiene è un integrale

multidimensionale che descrive la distribuzione a posteriori. La marginalizzazione diventa ancora più utile nel momento in cui la quantità di dati diminuisce rispetto al numero di parametri da stimare, ad esempio quando vengono utilizzati modelli molto complessi.

Perciò con l'analisi Bayesiana è necessario specificare un modello e una distribuzione a priori, quindi integrare il prodotto di queste due quantità su tutti i possibili valori dei parametri per determinare la probabilità a posteriori. Poiché le funzioni di likelihood implementate negli approcci Bayesiani per la stima di modelli evolutivisti sono troppo complesse da integrare in modo analitico, si usano i metodi MCMC per approssimare la distribuzione di probabilità a posteriori. I metodi MCMC consentono di campionare dalla distribuzione di probabilità a posteriori e calcolare la superficie di likelihood approssimata dal campione casuale ottenuto esplorando correttamente lo spazio di probabilità.

Queste nuove tecniche consentono quindi di combinare modelli stocastici complessi con metodi di simulazione allo scopo di inferire i diversi parametri che descrivono la storia demografica delle popolazioni.

2. SCOPO DELLO STUDIO

Gli scopi principali di questo studio sono quattro:

- a) quantificare i cambiamenti temporali nella diversità genetica, quindi prima e dopo la fase di sovrasfruttamento, utilizzando le statistiche descrittive classiche;
- b) descrivere la struttura di popolazione, identificando in questo modo le possibili unità indipendenti;
- c) inferire la popolazione effettiva short-term e stimare i parametri long-term (θ e tasso di crescita);
- d) effettuare test per l'identificazione di eventi di bottleneck genetici.

3. MATERIALI E METODI

Allo scopo di studiare la storia demografica del tonno rosso del Mediterraneo centro-ovest sono state raccolte le frequenze alleliche di 8 loci microsatelliti dalla collezione di campioni storici di tonno rosso di Massimo Sella collezionati tra il 1911 e il 1926 (mar Adriatico, sud Tirreno e Mediterraneo centrale) e di sei campioni di tonno contemporanei (mar Adriatico, sud Tirreno, mar Ligure, Sardegna, Algeria e mare di Alboran).(Figura 10,Tabella 1)

Nome del marcatore	Tipo di marcatore	Range di dimensioni alleliche
Tth-4	DINUCLEOTIDE	124-170
Tth-5	TETRANUCLEOTIDE	123-131
Tth-10	TETRANUCLEOTIDE	112-136
Tth-34	TETRANUCLEOTIDE	95-187
Tth 1-31	DINUCLEOTIDE	90-140
Tth 208	DINUCLEOTIDE	140-220
Tth 157	DINUCLEOTIDE	105-129
Tth 62	DINUCLEOTIDE	81-123

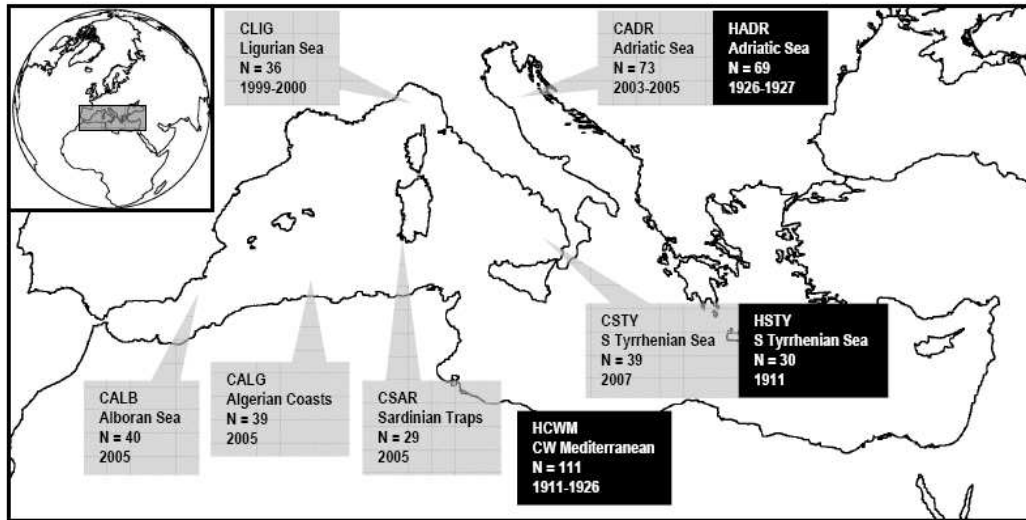
Tabella 1. Descrizione dei marcatori microsatelliti genotipizzati con numero di ripetizioni e dimensione.

3.1. Stima delle statistiche descrittive

Per studiare la composizione genetica di una popolazione è necessario fornire una descrizione quantitativa del pool genico di una popolazione. Questa stima viene fatta calcolando le frequenze genotipiche e le frequenze alleliche all'interno della popolazione. Le frequenze genotipiche per un dato locus rappresentano la frazione di individui con un dato genotipo sul totale di individui della popolazione, la somma delle frequenze dei diversi genotipi per quel locus deve essere uguale a 1. La frequenza del genotipo omozigote dominante AA può quindi essere indicata come $f(AA)$ e rappresenta il numero di individui con genotipo AA sul totale della popolazione, lo stesso vale per gli eterozigoti $f(aA)$ e gli omozigoti recessivi $f(aa)$.

La frequenza allelica è la misura della frequenza relativa di un allele ad un locus nella popolazione e di solito si esprime come proporzione o percentuale. In genetica di popolazioni, le frequenze alleliche descrivono la diversità genetica di una popolazione e la ricchezza del suo pool genico. La frequenza allelica può essere definita come la frazione o la percentuale di loci che l'allele occupa nella popolazione e può essere calcolata come proporzione del numero di individui che possiede un determinato allele sul numero totale di individui campionati 1), oppure dalle frequenze genotipiche 2):

a)



b)



c)



Figura 10. a) Località di campionamento degli esemplari storici (H, riquadri neri) e contemporanei (C, riquadri grigi) di *Thunnus thynnus* nel Mediterraneo centro-ovest. Per ognuno sono indicati l'acronimo, l'area geografica, la grandezza del campione e l'anno di campionamento. Per il campione HCWM non sono disponibili dati geografici. **b)** Tonni giganti catturati i primi del '900. **c)** Coda di un esemplare storico della collezione di Massimo Sella.

1) Frequenza allelica = numero di copie di un allele / totale degli alleli nella popolazione

2) $f(A) = 2 \times \text{numero omozigoti AA} + \text{eterozigoti Aa} / 2 \times \text{numero totale individui}$

$f(a) = 2 \times \text{numero omozigoti aa} + \text{eterozigoti Aa} / 2 \times \text{numero totale individui}$

Un'altra importante misura della diversità genetica di una popolazione è rappresentata dalla ricchezza allelica che misura il numero di alleli indipendentemente dalla dimensione della popolazione e perciò consente di confrontare accuratamente anche campioni di dimensioni diverse. Il metodo utilizza l'indice di rarefazione (El Mousadik and Petit 1996, Petit *et al.* 1998) e il suo principio è stimare il numero di alleli attesi in un sotto-campione di $2n$ geni, dato che sono stati campionati $2N$ geni ($N \geq n$). Il programma usato (vedere più avanti) considera che n sia fissato come il numero più piccolo di individui tipizzato per un locus in un campione.

La ricchezza allelica viene quindi calcolata come:

$$R_s = \sum_{i=1}^{ni} \left[1 - \frac{\binom{2N - Ni}{2n}}{\binom{2N}{2n}} \right]$$

dove N_i è il numero di alleli di tipo i tra i $2N$ geni. Ogni termine all'interno della sommatoria corrisponde alla probabilità di campionare l'allele i almeno una volta in un campione di dimensione $2n$.

Le frequenze alleliche, genotipiche e la ricchezza allelica sono state calcolate con il programma FSTAT versione 2.9.3.2 (Goudet 1995). FSTAT è un pacchetto di programmi che consentono di stimare ed effettuare test sulla diversità genetica e sulle statistiche che descrivono il differenziamento di popolazioni utilizzando markers genetici codominanti. Permette di calcolare diversi tipi di stimatori di diversità genetica e statistica F e di testare la loro significatività usando metodi di randomizzazione.

Uno dei principali parametri che di solito si stimano per ottenere informazione sulla diversità genetica di una popolazione è l'eterozigosi. Livelli elevati descrivono una popolazione che possiede una buona variabilità genetica, mentre bassi livelli di eterozigosi possono essere il segnale di una riduzione di variabilità dovuta ad un collo di bottiglia o ad altri fenomeni come la selezione. Di solito l'eterozigosi osservata nei campioni studiati viene confrontata con l'eterozigosi attesa sotto l'ipotesi di una popolazione in equilibrio di Hardy-Weinberg: se

L'eterozigosi osservata è più bassa di quella attesa potremmo essere di fronte ad un fenomeno di inbreeding, se al contrario è più alta si può ipotizzare il mescolamento tra due popolazioni precedentemente isolate. L'eterozigosi attesa (H_E) è calcolata sulla base dell'equilibrio di Hardy-Weinberg come:

$$H = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2 \right)$$

L'eterozigosi attesa (H_E) ed osservata (H_O), e il test di HW all'equilibrio sono stati calcolati con Arlequin 3.1 (Excoffier 2005), usando 100000 iterazioni delle catene di Markov e 1000 dememorisation steps. Il numero di dememorisation steps rappresenta il numero di iterazioni da effettuare prima di cominciare a confrontare le tabelle di probabilità delle ipotesi alternative con quelle osservate. Un migliaio di iterazioni sono sufficienti per ottenere una tabella indipendente da quelle osservate. Arlequin è un programma che integra diversi metodi di analisi di base ed avanzati di genetica di popolazioni, consente di calcolare frequenze alleliche e aploipiche, gli indici di diversità genetica classica, test di linkage disequilibrium, test di neutralità selettiva ed equilibrio demografico e di effettuare analisi di suddivisione di popolazione con il metodo dell'AMOVA. Il programma GENEPOP 3.1 (Raymond 1995) è stato usato per i test di significatività globale sui loci delle statistiche stimate (10000 dememorisation steps, 100 batch e 5000 iterazioni per ogni batch; i batch rappresentano una suddivisione del numero totale di permutazioni usate per stimare l'errore standard del P-value).

Gli indici di fissazione F_{ST} , F_{IS} e F_{IT} , anche definite statistiche F di Wright offrono un modo conveniente di rappresentare la struttura di una popolazione e la sua suddivisione in sottopopolazioni. I tre indici rappresentano la suddivisione in tre coefficienti della varianza allelica di un dato gene nell'intera popolazione, consentono cioè di misurare la deviazione delle frequenze di eterozigoti osservati da quelle attese secondo l'equilibrio di Hardy-Weinberg. Mentre F_{IS} e F_{IT} , rappresentano la correlazione tra due alleli prelevati da una popolazione in relazione rispettivamente ad una sottopopolazione e all'intera popolazione, F_{ST} misura la ripartizione della variabilità genetica tra sottopopolazioni; in altre parole, confronta la diversità genetica media presente all'interno delle sottopopolazioni con la diversità genetica della meta-popolazione.

I tre indici sono collegati dall'espressione:

$$F_{IS} = (F_{IT} - F_{ST}) / (1 - F_{ST})$$

Se la popolazione totale non presenta una suddivisione in sottopopolazioni avremo

$$F_{ST} = 0 \Rightarrow F_{IS} = F_{IT}.$$

La stima classica dell' F_{ST} di Wright si ottiene come

$$F_{ST} = \sigma_p^2 / p_i(1-p_i)$$

dove σ_p^2 è la varianza della frequenza allelica nelle sottopopolazioni, p_i è la frequenza allelica media fra sottopopolazioni e $p_i(1-p_i)$ è proporzionale all'eterozigosità attesa per l'intera popolazione e rappresenta il valore massimo che σ_p^2 può assumere. Perciò F_{ST} può essere interpretato come una misura percentuale della varianza genetica fra sottopopolazioni, rispetto al valore massimo osservabile date quelle frequenze alleliche (Wright 1969). La stima delle statistiche F di Wright è stata effettuata con Arlequin, mentre il calcolo dei valori globali di F_{ST} con FSTAT. La significatività dei valori di F_{IS} è stata calcolata effettuando 10000 permutazioni dei dati.

Durante la fase di genotipizzazione dei loci microsatelliti mutazioni a livello dei siti di legame dei primer possono non consentire l'amplificazione del corrispondente tratto di DNA. Questo comporta che l'allele non viene identificato e viene definito nullo, per cui individui eterozigoti vengono tipizzati come omozigoti. Questo può provocare una distorsione nella stima di parametri di genetica di popolazioni, quindi allo scopo di escludere l'effetto di alleli nulli sulla stima dell'indice di inincrocio, il contributo di ogni locus nella stima di F_{IS} è stato valutato usando il metodo non-parametrico di ricampionamento denominato jackknife. Questa procedura consiste nell'omettere, a turno, un locus e nel ricalcolare la statistica in esame, in questo caso F_{IS} , e confrontarla con quella globale per stabilire se uno dei loci ha provocato una distorsione nella stima dovuta, appunto, alla presenza di alleli nulli. La stima a coppie dell' F_{ST} è stata calcolata dal numero di alleli diversi e la distribuzione nulla di non differenziamento tra le sottopopolazioni è stata ottenuta con 10100 permutazioni. Con lo stesso programma è stata effettuata anche l'analisi della varianza molecolare (AMOVA, (Excoffier *et al.* 1992), in modo da quantificare la variabilità genetica tra i campioni sia attraverso lo spazio che il tempo (la significatività è stata valutata con 10100 test di permutazione). Questa analisi permette di definire la distribuzione della variabilità (espressa come percentuale di variabilità), all'interno della popolazione in esame, fra sottopopolazioni e fra individui, e quindi permette di determinare quantitativamente il grado di differenziazione tra popolazioni e testare statisticamente il livello di tali differenze. In particolare il programma attraverso l'analisi gerarchica della varianza divide la varianza totale nelle rispettive

componenti di covarianza dovute alle differenze intra-individuali, differenze inter-individuali e/o differenze inter-popolazioni. I componenti della covarianza vengono usati per stimare gli indici di fissazione di Wright. Poiché queste analisi implicano il confronto tra più di due test, cioè il confronto tra test multipli, non è possibile usare semplicemente α (probabilità di rifiutare l'ipotesi alternativa vera o errore di tipo I) come misura del tasso d'errore poiché esso aumenta all'aumentare del numero di confronti ed aumenta di conseguenza anche la probabilità di identificare dei falsi positivi. Per ovviare a questo problema è stata applicata la correzione sequenziale di Bonferroni per valutare la significatività dei confronti multipli (Rice 1989).

3.2. Struttura di popolazione

I genotipi individuali sono stati suddivisi in cluster utilizzando un approccio basato sui metodi Monte Carlo Markov Chain implementati nel programma STRUCTURE v2.2 (Pritchard *et al.* 2000). Il modello Bayesiano implementato in STRUCTURE valuta la presenza di Hardy Weinberg o linkage disequilibrium nel campione in esame e tenta di raggruppare le popolazioni in modo da minimizzare il disequilibrium e le deviazioni dalle frequenze alleliche di equilibrio all'interno di ogni gruppo. Il programma stima quindi la probabilità che un genotipo cada in ogni cluster così definito e calcola anche il coefficiente di membership, ossia la probabilità che ha un individuo di appartenere ad ogni cluster, consentendo di assegnarlo al gruppo con maggiore probabilità. Il modello scelto per queste analisi considera che gli individui possano aver avuto origine da una popolazione mescolata (admixture model), inoltre tiene conto della correlazione che esiste tra loci legati sullo stesso cromosoma (admixture linkage disequilibrium)(Falush *et al.* 2003). Sono state effettuate 10 repliche per ogni analisi con K che variava da 1 a 10 (K= numero di cluster in cui suddividere le popolazioni), con una fase di burn-in di 30000 step e 10000 iterazioni MCMC.

3.3. Stime demografiche da confronti temporali tra campioni

Per stimare la popolazione effettiva (N_e) sono stati utilizzati tre metodi diversi. Due di questi metodi sono chiamati 'metodi temporali' e consentono di stimare una N_e a breve termine (short-term) utilizzando i dati di campioni storici e contemporanei e riflette le fluttuazioni demografiche avvenute tra i due campionamenti.

Il metodo di Berthier *et al.* (2002), tramite il programma TM3, stima la N_e in base alla likelihood integrando un approccio basato sulla coalescenza. Questo metodo utilizza nell'analisi le genealogie (Figura 11) e non le frequenze geniche della popolazione come fanno programmi precedenti (Anderson *et al.* 2000), inoltre è possibile inserire una informazione a priori Bayesiana, ossia il valore massimo che può assumere la N_e ($N_{e\max}$).

Questo metodo assume che due campioni siano stati prelevati da una popolazione chiusa in due momenti diversi, che i due campioni siano indipendenti e che ognuno dei loci presente nei dati sia indipendente dagli altri. Per ottenere il valore di likelihood globale è quindi sufficiente moltiplicare la probabilità su tutti i campioni e loci. Usando il modello standard di coalescenza è possibile calcolare la probabilità di ottenere un determinato numero di eventi di coalescenza che dipendono da N_e . Nel caso di popolazioni che fluttuano nel tempo il valore di N_e stimato rappresenta la media armonica della popolazione effettiva stimata nell'intervallo di tempo relativo al campionamento (Wright 1938). Per ottenere la distribuzione degli eventi di coalescenza nell'intervallo e perciò la probabilità congiunta delle configurazioni del campione viene usato il metodo di Griffiths e Tavaré (Griffiths and Tavaré 1994). Il calcolo della probabilità congiunta su tutti i campioni e loci, quindi il campionamento dalla distribuzione a posteriori, viene effettuata con l'algoritmo Metropolis-Hastings. Il programma TM3 fornisce in output la moda del valore di N_e e l'intervallo credibile del 90%.

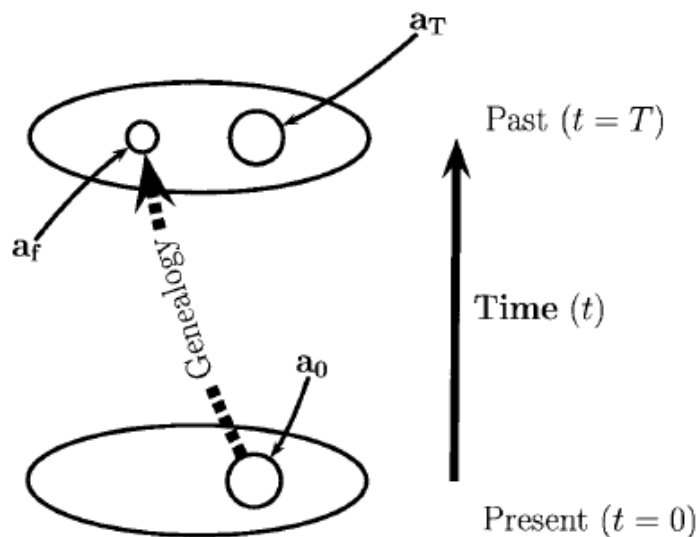


Figura 11. Modello usato dal nuovo stimatore (Berthier *et al.* 2002) basato sulla likelihood. Assume che i campioni siano presi da una popolazione chiusa in due momenti differenti. La popolazione è rappresentata dalla freccia tratteggiata. Seguendo le convenzioni dei modelli genealogici il tempo procede nel passato, con il campione più recente (a_0) preso al tempo $t = 0$ e quello precedente (a_f) al tempo $t = T$. a_0 si assume che abbia una genealogia descritta dal modello standard di coalescenza, e a_f rappresenta il fondatore.

L'altro metodo temporale è stato ideato da Wang (2001) e rappresenta un classico metodo basato sul calcolo della likelihood, dove le frequenze alleliche dei due campioni temporali e la relativa popolazione effettiva vengono stimati usando una matrice di transizione. Per poter calcolare la probabilità con loci multiallelici e rendere perciò più semplice la stima di probabilità multinomiale che ne deriva, Wang ha semplificato un allele a k-loci in k pseudoloci biallelici. A questo punto il calcolo della log-likelihood globale può essere approssimata mediante la somma di tutte le log-likelihood relative a tutti gli pseudoloci moltiplicata per il fattore $(k-1)/k$ che corregge per la dipendenza tra i loci. Questo metodo viene definito pseudo-likelihood perché consente di approssimare la probabilità congiunta con il prodotto delle probabilità marginali. Il metodo di Wang e il programma MLNE1 che lo applica, forniscono in output la media armonica di N_e e l'intervallo di confidenza al 95%.

La stima della short-term N_e è stata effettuata soltanto per i due campioni di cui avevamo esattamente la replica temporale (mar Adriatico e Sud Tirreno). Il numero di generazioni tra i due campionamenti è stato stimato considerando 4 anni (Fromentin and Powers 2005) come tempo di generazione per il tonno rosso che rappresentano perciò 24 generazioni per il campione del Tirreno e 20 per l'Adriatico. Entrambe i metodi assumono un modello con generazioni discrete, tuttavia il tonno rosso possiede generazioni sovrapposte come la maggior parte dei pesci, ma il numero di generazioni che separano i due campioni temporali (~20) sembrano essere sufficienti per minimizzare qualsiasi distorsione dovuta alla struttura di età (Waples *et al.* 2007). Con il metodo di Berthier sono state effettuate alcune repliche dell'analisi, ognuna costituita da 300000 iterazioni MCMC fissando la $N_{e\max}$ a 5000, 10000 e 50000 per verificare che cambiando il range della priori non cambiasse anche la stima (anche per il metodo di Wang). La convergenza per il programma TM3 è stata verificata con il programma Tracer v1.4 (Rambaut and Drummond 2004).

3.4. Stime demografiche da singoli campioni

Oltre ai due metodi già descritti, per stimare la N_e è stato usato un terzo metodo descritto da Kuhner 2006. I pattern di variabilità genetica in una popolazione vengono modellati da interazioni a lungo termine delle forze evolutive (selezione, migrazione deriva e mutazione). Nella situazione più semplice di una popolazione isolata con N_e costante, la variabilità genetica di marcatori neutrali è frutto del contributo di mutazione e deriva. Nella situazione all'equilibrio mutazione-deriva la diversità genetica è costante ed è descritta dal parametro di popolazione $\theta=4N_e\mu$, dove μ rappresenta il tasso di mutazione per generazione e locus. La

variabilità genetica di un campione di individui caratterizzati mediante marker neutrali può quindi essere utilizzata per inferire θ e ottenere, nel caso si conosca il tasso di mutazione, il relativo valore di N_e ; se μ non è noto è possibile confrontare i valori di θ stimati per popolazioni diverse.

Il metodo di Kuhner (Kuhner *et al.* 1998, Kuhner 2006) è definito long-term perché stima l'effetto a lungo termine delle forze evolutive tramite la stima di θ . Il programma che applica questo metodo, LAMARC, permette di effettuare sia analisi mediante likelihood che mediante un approccio Bayesiano. Questo secondo metodo, usato per le nostre analisi, è rappresentato da un campionatore Bayesiano che sceglie a caso sia le genealogie simulate mediante coalescenza che i parametri dalla distribuzione a priori. I valori dei parametri che vengono visitati più spesso dal campionatore rappresentano la distribuzione a posteriori dei parametri. Il metodo Bayesiano effettua degli step di cambiamento delle genealogie utilizzando i valori correnti dei parametri come 'driving values'. Inoltre introduce anche uno step di cambiamento dei parametri; l'output di una catena non è infatti rappresentato da un campione di genealogie come nell'analisi di likelihood, ma è rappresentato da un campione di combinazioni di parametri che sono stati visitati dalla catena. Nel momento in cui viene proposto un cambiamento di parametro, viene scelto un parametro a caso tra quelli presenti nel modello e il suo valore viene sorteggiato in modo aleatorio dalla distribuzione a priori. La probabilità della genealogia corrente basata sui parametri $[\text{Prob}(G|P)]$ viene calcolata sia per il set di parametri vecchi che per quelli nuovi. I nuovi valori vengono accettati se una frazione uniforme di valori casuali ha una probabilità più bassa di $U < (P(G|P_{\text{new}}))/(P(G|P_{\text{old}}))$, altrimenti vengono tenuti i parametri vecchi.

Per ogni popolazione sono stati stimati sia θ che il tasso di crescita (g) effettuando due repliche dell'analisi sotto il modello di crescita esponenziale; per ogni replica sono state create un miliardo di genealogie da cui sono state campionate una genealogia ogni 10000, con un burn-in di 1000 iterazioni. La mutazione è stata modellata secondo l'approssimazione Brownian-motion del modello di mutazione stepwise (Beerli and Felsenstein 2001). Il modello Brownian-motion considera il processo di mutazione come un percorso continuo e casuale dal punto di partenza, questa approssimazione matematica rende il calcolo molto più veloce rispetto al classico modello di mutazione stepwise per loci microsatelliti che assume che ogni evento di mutazione aumenti o diminuisca il numero di ripetizioni di una singola ripetizione alla volta.

I valori di g (misurati in $1/\mu$ generazioni) e di θ sono stati ottenuti da due distribuzioni a priori uniformi ($1^{-10} \leq \theta \leq 100$; $-500 \leq g \leq 100$). Anche in questo caso il programma Tracer v1.4 (Rambaut and Drummond 2004) è stato usato per controllare che le catene fossero arrivate a convergenza.

3.5. Verifica degli effetti di crisi demografiche

Come ultima analisi sono stati eseguiti due test per determinare se le popolazioni hanno subito un bottleneck che ha provocato una diminuzione della variabilità genetica.

Il primo metodo presente nel programma BOTTLENECK v.1.2.02 (Cornuet and Luikart 1996, Piry *et al.* 1999), assume che in seguito ad un collo di bottiglia il numero di alleli presenti in una popolazione diminuisca più velocemente dell'eterozigosi, viene quindi confrontata l'eterozigosi attesa (H_E) con l'eterozigosi all'equilibrio stimata dal numero di alleli della popolazione (H_{EQ}). In pratica è un test che consente di stabilire se il campione in esame presenta un deficit di alleli, quindi un eccesso di eterozigosi e perciò ha probabilmente subito un collo di bottiglia.

Mediante simulazioni Cornuet *et al.* (1996) hanno esteso il tipo di analisi consentite. Grazie al loro lavoro, oltre al modello di mutazione IAM (*Infinite Alleles Model*) è possibile considerare nelle analisi il modello SMM (*Stepwise Mutation Model*) e TPM (*Two Phase Model*). Il modello IAM considera tutti gli alleli equivalenti, perciò la probabilità di osservare una mutazione da uno qualsiasi di questi alleli ad un altro qualsiasi è la stessa. I modelli SMM e TPM descrivono in modo più preciso il tipo di mutazione caratteristica dei loci microsatelliti: il modello SMM consente solo mutazioni one-step (una sola ripetizione viene aggiunta o deleta ad ogni evento mutazionale); il modello TPM tiene conto della varianza presente sia nel numero di mutazioni non one-step che nella lunghezza dell'unità che viene aggiunta o eliminata ad ogni evento mutazionale. In pratica mediante il modello standard di coalescenza si ottengono per simulazione i valori di H_E . Per ogni albero simulato vengono sovrapposte le mutazioni in base al modello mutazionale scelto. Ripetendo questo processo molte volte si ottiene la distribuzione all'equilibrio di H_E condizionato il numero di alleli presenti nel campione e la sua dimensione. A questo punto possono essere usati diversi metodi statistici per stabilire se lo scostamento dall'equilibrio mutazione-deriva è significativo. Nel nostro caso il test scelto è il test di Wilcoxon a una e due code calcolato per ogni tasso di mutazione (per il modello TPM è stata utilizzata una percentuale del 5% di

eventi multistep e 95% one step e una varianza nella lunghezza dell'unità che viene aggiunta o eliminata ad ogni evento mutazionale di 12 come in Piry *et al.* 1999).

Il secondo metodo, disponibile nel programma M_P_Val (Garza and Williamson 2001), prevede il calcolo del rapporto M tra il numero di alleli (k) e il range della grandezza allelica (r). Quando le dimensioni di una popolazione si riducono, l'effetto della deriva genetica viene amplificato con conseguente perdita di alleli. Tuttavia qualsiasi allele perso contribuisce alla diminuzione del numero di alleli k, mentre solo quelli più grandi o più piccoli ridurranno il range r, quindi k si riduce più velocemente di r e nelle popolazioni in bottleneck il rapporto $M=k/r$ sarà più basso rispetto alle popolazioni all'equilibrio. In generale la probabilità che un allele venga perso a causa della deriva è collegata alla sua frequenza e gli alleli rari vengono persi più facilmente e contribuiscono a diminuire k, ma non hanno lo stesso effetto su r, a meno che non siano gli alleli di minima o massima lunghezza. Il programma permette di considerare diversi scenari per il modello mutazionale TPM, in cui scegliere valori diversi per p_s (frequenza delle mutazioni multistep) e Δg (dimensione media delle mutazioni non one-step). Sono stati esaminati diversi scenari mutazionali (come raccomandato da Garza *et al.* 2001) e sono stati scelti dei valori conservativi per i due parametri $p_s=0.1$ e $\Delta g=3.5$. Con M_P_Val è possibile anche calcolare un valore di M definito M critico (M_{crit}) che è il risultato di 10000 simulazioni di popolazioni all'equilibrio di cui si calcola il valore M. M_{crit} rappresenta il valore in cui il 95% delle popolazioni simulate all'equilibrio ha $M > M_{crit}$. Per ogni scenario considerato sono stati utilizzati 4 diversi valori di θ : i valori ottenuti dalle stime short-term di N_e (usando $\mu=10^{-4}$ (Waldick *et al.* 2002), i valori long-term stimati con LAMARC e due valori comunemente osservati nelle popolazioni naturali ($\theta=0.5$ e $\theta=10$ (Guinand and Scribner 2003).

4. RISULTATI

4.1. Analisi delle statistiche descrittive

L'analisi delle statistiche classiche è riassunta nella Tabella 1. Il numero medio di alleli varia da 8.6 (CSAR) a 13.13 (CADR) nei campioni contemporanei, in quelli storici va da 9.5 per HSTY e HCWM a 10.88 per HADR. Il valore di allelic richness nei campioni contemporanei è compreso tra 8.48 e 10, mentre in quelli storici tra 7.13 e 9.34. Il campione HCWM mostra una distribuzione genotipica insolita, con un forte eccesso di eterozigoti a molti loci (Tabella 2), per esempio l'allele Tth157*2 è presente in 85 copie e l'allele Tth10*2 in 60 copie, ma non sono presenti alleli omozigoti *2*2 in nessuno dei due loci. Poiché non è possibile spiegare in modo logico questo risultato, il campione HCWM non è stato utilizzato nelle analisi successive.

Tutti i campioni tranne CSTY mostrano una deviazione dall'equilibrio di Hardy Weinberg delle frequenze alleliche sui loci, tutte dovute ad un deficit di eterozigoti. La maggior parte dei coefficienti di inbreeding (F_{IS}) sono positivi e significativi (Tabella 3). La presenza di alleli nulli è stata esclusa effettuando un'analisi di jackknife sui loci e ricalcolando ogni volta il coefficiente di inbreeding. Le uniche due eccezioni sono il campione CALB, in cui la deviazione dall'equilibrio non è significativa, e CSTY in cui è presente un eccesso di eterozigoti non significativo. In generale le differenze nel numero di alleli, ricchezza allelica e diversità genica tra campioni contemporanei e storici non sono risultate statisticamente significative (test U Mann-Whitney).

Campione	Locus							
	Tth4	Tth5	Tth10	Tth34	Tth1-31	Tth208	Tth157	Tth62
CSAR								
n	29	28	29	29	29	29	29	28
a	12	3	2	11	10	17	4	10
aR	11.58	3	2	10.85	9.86	16.64	4	9.89
aS	124-168	123-131	116-120	103-147	92-126	140-208	117-125	83-109
He	0.82	0.49	0.41	0.81	0.83	0.92	0.45	0.85
Ho	0.90	0.32	0.34	0.79	0.72	0.76	0.41	0.82
HW	0.44	0.09	0.64	0.22	0.77	0.00	0.29	0.04
CADR								
n	73	71	73	73	73	71	73	71
a	16	3	5	17	18	26	7	13
aR	12.90	2.91	3.64	12.99	14.04	18.25	5.34	10.17
aS	130-170	123-131	112-136	103-183	90-128	140-206	117-129	83-113
He	0.82	0.47	0.51	0.83	0.89	0.92	0.66	0.84
Ho	0.75	0.42	0.55	0.73	0.84	0.70	0.51	0.79
HW	0.05	0.13	0.00	0.02	0.12	0.00	0.00	0.51
CLIG								
n	35	35	36	36	36	36	36	36
a	11	3	2	14	13	18	6	13
aR	10.03	3	2	12.48	12.19	16.34	5.75	11.69
aS	136-168	123-131	116-120	103-163	92-130	140-206	119-129	85-123
He	0.81	0.54	0.48	0.84	0.90	0.92	0.70	0.87
Ho	0.83	0.43	0.44	0.86	0.86	0.69	0.61	0.78
HW	0.64	0.11	0.73	0.51	0.39	0.01	0.03	0.02
CSTY								
n	39	39	39	39	39	38	38	39
a	12	3	2	12	17	19	6	12
aR	10.86	2.69	2	11.38	15.03	17.30	5.13	11.02
aS	130-170	123-131	116-120	107-187	92-140	140-206	117-127	83-115
He	0.85	0.41	0.43	0.82	0.92	0.91	0.48	0.84
Ho	0.77	0.46	0.56	0.92	0.92	0.87	0.58	0.85
HW	0.24	0.17	0.07	0.00	0.15	0.00	0.05	0.15
CALG								
n	39	39	39	38	38	39	39	39
a	13	3	2	17	15	18	6	11
aR	11.90	3	2	14.07	12.87	15.96	5.79	10.26
aS	136-168	123-131	116-120	103-187	90-132	140-216	119-129	83-115
He	0.87	0.53	0.49	0.75	0.87	0.91	0.62	0.82
Ho	0.79	0.38	0.54	0.74	0.84	0.69	0.62	0.90
HW	0.11	0.07	0.74	0.78	0.07	0.00	0.87	0.73
CALB								
n	40	40	40	40	40	38	40	40
a	11	3	2	13	13	18	4	10
aR	10.03	2.99	2	11.87	11.81	16.55	4	8.81
aS	138-170	123-131	116-120	103-155	92-138	140-194	119-125	83-111
He	0.80	0.49	0.51	0.83	0.90	0.93	0.50	0.74
Ho	0.70	0.60	0.58	0.80	0.80	0.87	0.40	0.65
HW	0.08	0.01	0.52	0.98	0.00	0.18	0.01	0.00
HSTY								
n	30	30	30	30	29	27	30	30
a	10	3	2	10	9	19	8	15
aR	9.60	2.90	2	9.88	8.92	19	7.78	14.66
aS	136-170	123-131	116-120	95-151	92-126	140-202	105-125	83-117

He	0.81	0.41	0.44	0.85	0.85	0.93	0.68	0.88
Ho	0.50	0.40	0.43	0.83	0.79	0.63	0.67	0.67
HW	0.00	1.00	1.00	0.01	0.00	0.00	0.49	0.00
HADR								
n	69	69	68	69	68	68	67	68
a	11	3	2	14	16	22	6	13
aR	9.13	2.92	2	10.90	12.09	16.39	5.43	11.15
aS	124-170	123-131	116-120	99-155	90-128	140-196	117-127	81-115
He	0.83	0.41	0.46	0.85	0.88	0.88	0.70	0.83
Ho	0.68	0.46	0.47	0.71	0.72	0.68	0.60	0.78
HW	0.00	0.63	1.00	0.05	0.00	0.00	0.00	0.24
HCWM								
n	107	105	107	107	107	107	107	107
a	8	3	2	11	16	24	5	7
aR	5.15	2.99	2	7.86	11.65	16.66	4.93	5.80
aS	134-148	123-131	116-120	103-147	90-130	140-220	117-125	91-103
He	0.65	0.57	0.41	0.62	0.86	0.90	0.73	0.70
Ho	0.36	0.87	0.56	0.62	0.76	0.81	0.96	0.81
HW	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Tabella 2.- Variabilità delle statistiche descrittive calcolate per gli 8 loci microsatelliti e i nove campioni. \bar{a} , numero medio di alleli; $\bar{a}R$, ricchezza allelica media; H_e , eterozigosità attesa media; H_o , eterozigosità media osservata; HW, test dell'equilibrio di Hardy-Weinberg.

Campione	Diversità genetica				P_{HW}	F_{IS}	P_{FIS}
	\bar{a}	$\bar{a}R^*$	H_e^*	H_o^*			
HADR	10.9	8.7	0.73	0.64	0.0000	0.119	0.0000
HSTY	9.5	9.3	0.73	0.62	0.0000	0.140	0.0001
CADR	13.1	10.0	0.74	0.66	0.0000	0.099	0.0000
CSTY	10.4	9.4	0.71	0.74	NS	-0.057	NS
CLIG	10.0	9.2	0.76	0.69	0.0000	0.087	0.0029
CSAR	8.6	8.5	0.70	0.63	0.0004	0.088	0.0094
CALG	10.6	9.5	0.73	0.69	0.0012	0.054	0.0375
CALB	9.3	8.5	0.71	0.67	0.0028	0.046	NS

Tabella3.- Diversità genetica, test dell'equilibrio di Hardy-Weinberg e stime di F_{IS} per i campioni di *Thunnus thynnus*. \bar{a} , numero medio di alleli; $\bar{a}R$, ricchezza allelica media; H_e , eterozigosità attesa media; H_o , eterozigosità media osservata; P_{HW} , P-value del test di H-W; F_{IS} , coefficiente di inbreeding; P_{FIS} , P-value dei valori di P_{FIS} calcolati mediante 10100 permutazioni. NS, non significativo. *Tutti i test Us Mann-Whitney's tra campioni storici e contemporanei erano non significativi.

4.2. Analisi di struttura di popolazione

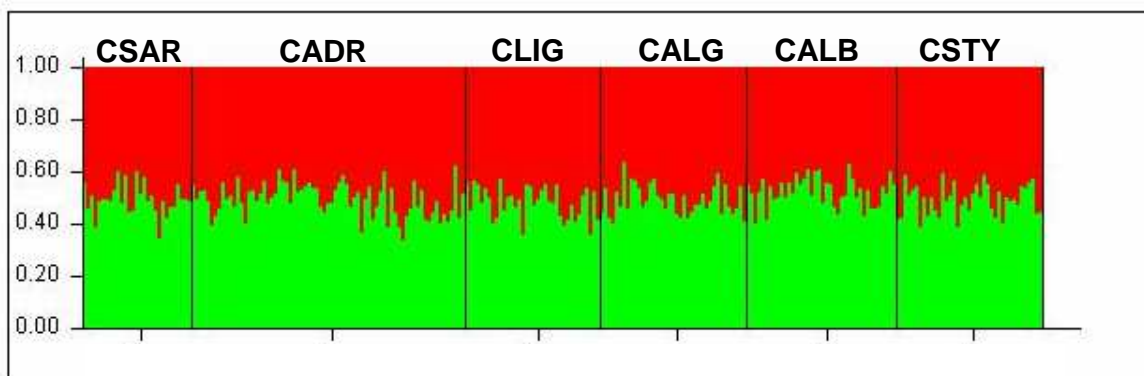
La maggior parte degli F_{ST} calcolati a coppie tra i campioni contemporanei risultano significativamente maggiori di 0, con l'eccezione dei confronti CADR-CLIG e CSTY-CALG (Tabella 4). In realtà, in seguito a correzione sequenziale di Bonferroni per test multipli, solo il confronto CALB-CSTY non risulta significativo. Il differenziamento genetico globale è molto simile tra i campioni contemporanei ($F_{ST}=0.026$, $P<0.0001$) e tra quelli storici ($F_{ST}=0.020$, $P<0.0001$). La presenza di un segnale significativo di differenziamento tra tutti i campioni è stato confermato anche utilizzando l'Analisi della Varianza Molecolare (AMOVA: 4.47% della varianza totale $P<0.0001$), sebbene le differenze siano significative solo all'interno dei gruppi, ossia tra i campioni contemporanei e tra quelli storici (3.94%, $P<0.0001$) e non nel confronto globale tra contemporanei e storici (0.83%, $P=0.225$).

Le analisi effettuate con il programma STRUCTURE non hanno permesso di evidenziare il segnale di struttura individuato con l'analisi degli F_{ST} e con l'AMOVA. $K=1$ risulta infatti il numero di cluster più probabile, e per $K>1$ la probabilità di assegnamento a posteriori di ogni individuo è esattamente la stessa per ogni cluster, quindi molto probabilmente il programma non è in grado di individuare il segnale, forse a causa del basso segnale di struttura, del numero di loci non sufficienti o della grandezza del campione non adeguata (Grafico1)

	HADR	HSTY	CADR	CSTY	CLIG	CSAR	CALG
HSTY	0.020						
CADR	0.015	0.016					
CSTY	0.022	0.015	0.015				
CLIG	0.019	0.017	0.005*	0.018			
CSAR	0.057	0.081	0.045	0.088	0.049		
CALG	0.024	0.020	0.016	0.004*	0.016	0.082	
CALB	0.026	0.019	0.015	0.011**	0.019	0.089	0.015

Tabella 4.- Risultato dell'analisi a coppie degli F_{ST} . La significatività è stata testata mediante 10100 permutazioni. * valori non significativi al 5%. ** valori non significativi dopo correzione di Bonferroni ($\alpha = 0.0025$).

a)



b)

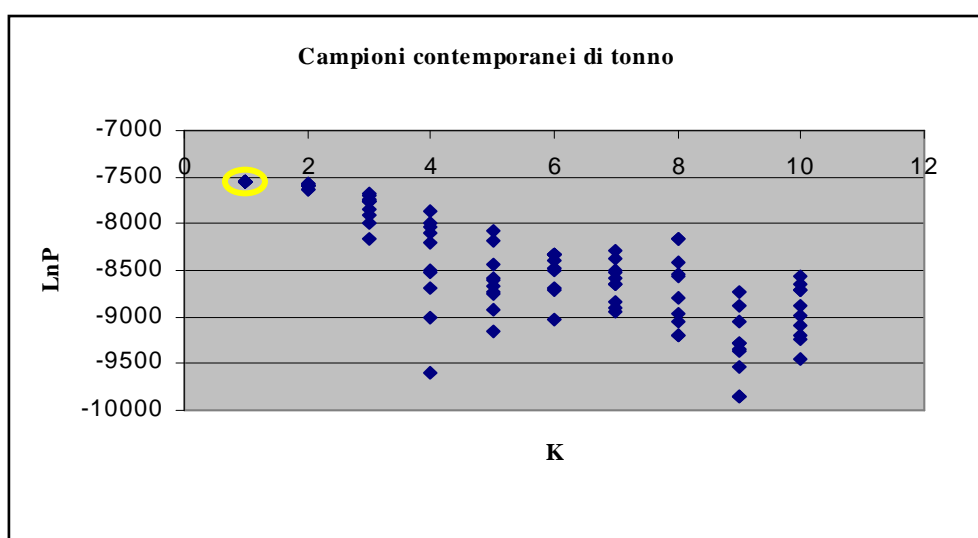


Grafico 1. Risultato dell'analisi di STRUCTURE per i campioni moderni. **a)** Proporzioni dell'appartenenza di ogni individuo di *Thunnus thynnus* considerando due cluster genetici (Q). **b)** Valori di likelihood ottenuti dall'analisi per ogni valore di K nelle 10 repliche effettuate. Il programma individua un solo cluster come più probabile (indicato dal cerchio giallo)

4.3. Inferenza stime demografiche

I valori short-term della dimensione effettiva di popolazione sono stati stimati con i due programmi descritti, MLNE1 (Wang 2001) e TM3 (Berthier *et al.* 2002) confrontando campioni storici e contemporanei raccolti nella stessa area geografica. Le stime ottenute e l'intervallo di confidenza al 95% (CI) sono coerenti, e forniscono un valore di N_e di circa 10^3 . Il valore di N_e stimato per il campione dell'Adriatico (MLNE: 1473, CI = 838-3610; TM3: 1260, CI = 696-2892), sembra essere più elevato rispetto al sud Tirreno (MLNE: 634, CI = 375-1354; TM3: 815, CI = 440-2124). L'analisi con il metodo basato sulla coalescenza TM3

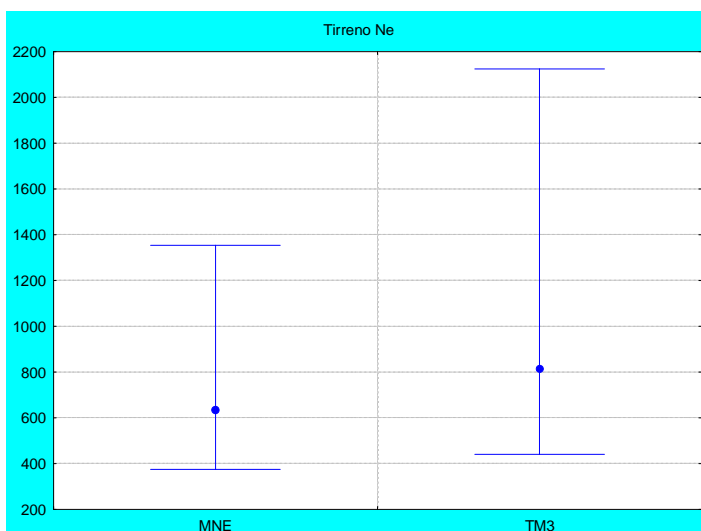
fornisce per il campione del sud Tirreno un valore leggermente superiore rispetto al metodo basato sulla likelihood MLNE, ma gli intervalli di confidenza sono comunque ampiamente sovrapposti. (Grafico 2)

Inoltre stime long-term di N_e sono state inferite dai livelli di diversità genetica presenti in ogni campione utilizzando l'analisi bayesiana delle genealogie basata sulla teoria della coalescenza (LAMARC, Tabella 5)(Kuhner 2006). Con questo metodo si sono ottenuti valori di $\theta (=4N_e\mu)$ nel range 21.9 (in CALB) – 39.9 (in HSTY), cioè dello stesso ordine di grandezza tra i campioni storici e contemporanei. Per ottenere i valori di N_e dalle stime di θ , è stato usato $\mu=10^{-4}$ (che rappresenta un valore intermedio tra i due ampiamente utilizzati, 10^{-3} - 10^{-5} (Waldick *et al.* 2002); i valori long-term così ottenuti sono molto maggiori dei valori short-term, variano infatti da 52070 a 99670. Anche utilizzando un valore estremamente elevato di tasso di mutazione (10^{-3}) come suggerito da Fraser *et al.* (2007), le stime long-term sono circa dieci volte più elevate di quelle short-term (Fraser *et al.* 2007).

Con lo stesso metodo è stato stimato anche il tasso di crescita (Tabella 5). Le stime puntuali ottenute sono tutte negative, per tutti i campioni sia storici che contemporanei, tuttavia gli intervalli di confidenza sono centrati sullo zero suggerendo che la popolazione possa essere stazionaria dal punto di vista demografico.

STY

MNe2				
PARAMETRO	STIMA	+95%CI	-95%CI	
Ne	634	1354	375	
TM3				
Ne	815	2124	440	



ADR

MNe2

PARAMETRO	STIMA	+95%CI	-95%CI
Ne	1473	3610	838

TM3

Ne	1260	2892	696
----	------	------	-----

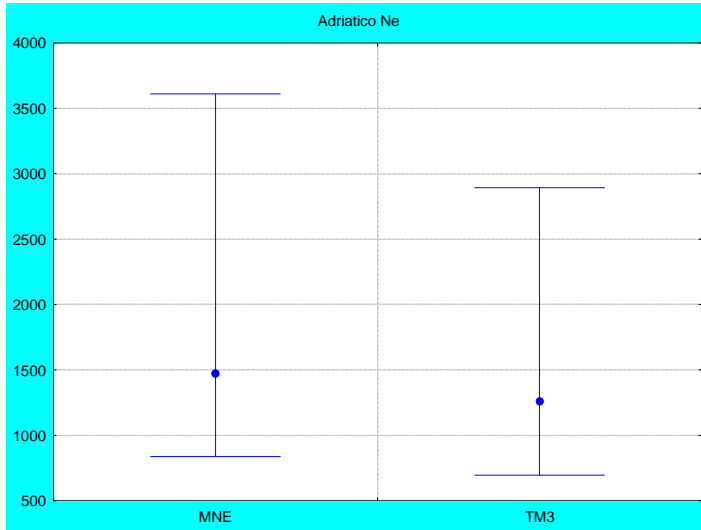


Grafico 2. Stime della popolazione effettiva (N_e) ottenute con i metodi temporali di Wang (2001) e Berthier *et al.* (2002). STY rappresenta la stima per i due campioni temporali del sud Tirreno e ADR per l'Adriatico.

Campione	θ		N_e	Tasso di crescita	
	Stima	95% CI		Stima (g/μ)	95% CI
HADR	29.7	21.7 - 40.4	74170	-0.106	-0.134 , 0.113
HSTY	39.9	18.3 - 44.0	99670	-0.011	-0.026 , 0.197
CADR	35.7	27.0 - 47.0	89250	-0.094	-0.100 , 0.130
CSTY	28.0	24.0 - 45.4	70000	-0.147	-0.152 , 0.047
CLIG	31.3	22.0 - 45.0	78300	-0.025	-0.043 , 0.147
CSAR	20.8	17.0 - 38.2	52070	-0.111	-0.138 , 0.064
CALG	29.3	20.6 - 42.0	73120	-0.001	-0.021 , 0.217
CALB	21.9	16.7 - 34.4	54750	-0.017	-0.041 , 0.198

Tabella 5.- Stime long-term di θ , N_e e del tasso di crescita nei campioni di tonno rosso del Mediterraneo. La N_e è ottenuta dai valori di θ usando il valore medio del tasso di mutazione per i microsatelliti ($\mu = 10^{-4}$). CI= Intervallo Credibile.

4.4. Identificazione dei segnali di crisi demografica

I due metodi usati per determinare se la popolazione ha subito un collo di bottiglia (BOTTLENECK(Piry *et al.* 1999) e M_P_Val(Garza and Williamson 2001)), si basano sullo stesso tipo di assunzione, ossia che gli alleli rari vengono persi con maggiore probabilità rispetto agli alleli comuni. Questo provoca un declino più rapido del numero di alleli rispetto

all' eterozigosità (Cornuet and Luikart 1996) e al range della dimensione allelica dei microsatelliti (Garza and Williamson 2001). Per testare quindi la presenza di un bottleneck demografico ho impiegato i due metodi descritti per confrontare i risultati ottenuti. Per l'approccio sviluppato in BOTTLENECK è necessario definire il modello mutazionale e come detto sono stati scelti tre modelli, cioè il modello degli alleli infiniti (IAM), quello delle mutazioni stepwise single-step (SMM) e il multi-step stepwise (TPM). Con il modello IAM tutti i campioni (Tabella 6) tranne CALG presentano un eccesso di eterozigosi, che rappresenterebbe un segnale di bottleneck. Tuttavia questo modello mutazionale non rappresenta in modo preciso il meccanismo di mutazione che caratterizza i loci microsatelliti (Spong and Hellborg 2002). Con gli altri modelli mutazionali i risultati sono in effetti radicalmente diversi. Dieci dei sedici test effettuati (due per campione), presentano risultati significativi, ma il segnale individuato è un deficit di eterozigosi che, se preso alla lettera, descriverebbe una popolazione in espansione. I risultati dei campioni CLIG e CALB non sono significativi sia con il modello SMM che TPM, mentre quelli di HSTY e CALG non sono significativi sotto il modello TPM.

Il test di bottleneck basato sulla stima del rapporto M (Garza and Williamson 2001) richiede la definizione a priori di θ . Per l'analisi sono stati usati tre valori di θ (Tabella 6), quelli stimati per ogni campione con l'analisi long-term e due valori spesso riscontrati nelle popolazioni naturali di pesci, 0.5 e 10 (Guinand and Scribner 2003). Per i campioni CADR e CSTY era inoltre disponibile la stima short-term di N_e con cui si è ottenuto il valore di θ ($\mu=10^{-4}$). Il valore per il campione CADR è esattamente 0.5, perciò per questo campione abbiamo una stima in meno rispetto agli altri campioni. Questo metodo si è dimostrato molto sensibile alla scelta del valore di θ . Tutti i valori stimati di M sono più bassi di 0.9 che rappresenta il valore atteso per una popolazione di dimensione costante, ma solo 7 dei 25 test sono significativi, e in nessun caso è stato possibile individuare un segnale di bottleneck indipendente dal valore di θ considerato. Quando si utilizza il valore 0.5, molto basso, sia campioni storici che contemporanei mostrano dei valori di M più bassi del valore critico M_{crit} , eccetto HADR e CADR. Il campione CSTY mostra un segnale di bottleneck anche con il valore short-term 0.27. Al contrario utilizzando i valori più elevati di θ non si riscontra nessuna evidenza di collo di bottiglia.

BOTTLENECK (Cornuet <i>et al.</i> 1996)							M ratio (Garza 2001)		
Campione	IAM		SMM		TPM		θ	M/M _{crit}	P-value
	P ₂	P ₁	P ₂	P ₁	P ₂	P ₁			
HADR	0.0078	0.0039 (E)	0.0078	0.0039 (D)	0.0117	0.0059 (D)	29.7	0.82/0.65	0.465
							0.5	0.82/0.75	0.056
							10	0.82/0.70	0.278
HSTY	0.0117	0.0059 (E)	0.0547	0.0273 (D)	NS	NS	39.9	0.73/0.60	0.490
							0.5	0.73/0.79	0.004
							10	0.73/0.66	0.095
CADR	0.0117	0.0059 (E)	0.0039	0.0020(D)	0.0039	0.0020 (D)	35.0	0.84/0.65	0.634
							0.5*	0.84/0.75	0.090
							10	0.84/0.70	0.392
CSTY	0.0742	0.0371 (E)	0.0391	0.0195 (D)	0.0742	0.0371(D)	0.27*	0.76/0.81	0.006
							28.0	0.76/0.65	0.287
							0.5	0.76/0.79	0.011
CLIG	0.0039	0.0020 (E)	NS	NS	NS	NS	31.3	0.79/0.64	0.660
							0.5	0.79/0.79	0.030
							10	0.79/0.67	0.300
CSAR	0.0117	0.0059 (E)	0.0391	0.0195 (D)	0.0742	0.0371(D)	20.8	0.75/0.63	0.320
							0.5	0.75/0.79	0.008
							10	0.75/0.66	0.148
CALG	NS	NS	0.0391	0.0195 (D)	NS	NS	29.3	0.79/0.73	0.523
							0.5	0.79/0.80	0.024
							10	0.79/0.68	0.239
CALB	0.0078	0.0039 (E)	NS	NS	NS	NS	21.9	0.80/0.75	0.521
							0.5	0.80/0.79	0.030
							10	0.80/0.68	0.317

Tabella 6.- Risultati delle analisi di bottleneck.

Nella analisi con BOTTLENECK (Cornuet *et al.* 1996) IAM è l'Infinite Allele Model (Kimura and Crow 1964); SMM è lo Stepwise Mutation Model (Ohta and Kimura 1973); TPM è il Two-Phase Model con percentuale di eventi multistep=5% e varianza=12 (vedere Piry *et al.* (1999)). P₂ rappresenta il P-value del test di Wilcoxon a due code per eccesso/deficit di eterozigotità; P₁ il P-value del test di Wilcoxon a una coda (E-eccesso, D-deficit). Nell'analisi del rapporto M (Garza *et al.* 2001), θ ($= 4N_e\mu$) è una stima della popolazione effettiva, M è il rapporto medio osservato tra il numero di alleli (k) e il range nella dimensione allelica (r) sui loci, M_{crit} è il valore critico di M stimato attraverso 10000 simulazioni ottenute con gli stessi parametri dei dati osservati, ma assumendo una popolazione all'equilibrio. Per ogni campione si sono effettuate stime di M usando tre valori di θ : stime long-term di LAMARC (prima riga), $\theta=0.5$ (seconda riga) e $\theta=10$ (terza riga); per i campioni CADR e CSTY è presente anche una stima effettuata con valori di θ short-term ottenuta dai metodi temporali (indicati con asterisco). I valori significativi sono indicati in grassetto.

5. DISCUSSIONE

La genetica applicata alla pesca, ovvero l'applicazione dei principi e metodi della genetica alla biologia della pesca e alla sua gestione, ha risollevato l'interesse verso tutti quei fattori su cui sono basate le dinamiche e le capacità di ripresa delle specie sfruttate. Argomenti quali per esempio la connettività tra le popolazioni marine (Cowen *et al.* 2006, Trembl *et al.* 2008), il differenziamento delle popolazioni su scala temporale e spaziale (Ruzzante *et al.* 2006), la dimensione della popolazione effettiva (Hauser *et al.* 2002, Waples and Yokota 2007), l'evoluzione indotta dalla pesca (Marshall and Browman 2007) e l'analisi della variabilità adattativa nelle popolazioni selvatiche (Conover *et al.* 2006), non solo ampliano le nostre conoscenze relative ai meccanismi che descrivono l'abbondanza dei pesci e la loro distribuzione, ma contribuiscono concettualmente alla teoria ecologica ed evolutiva. Gli approcci classici utilizzati nella pesca si concentrano sui fattori che provocano dei cambiamenti demografici a breve termine nelle popolazioni (cambiamenti 'quantitativi'), al contrario gli approcci genetici esaminano l'entità in cui i cambiamenti nella composizione delle popolazioni (cambiamenti 'qualitativi') influenzano sia alterazioni a breve termine dei tratti fenotipici sia la risposta a lungo termine alle perturbazioni naturali e antropogenetiche (Frankham 2005). Sebbene fosse già ampiamente riconosciuto che le dinamiche di popolazione e la struttura genetica sono collegate in modo imprescindibile, la biologia della pesca tradizionale fa scarso riferimento ai fattori genetici nell'ecologia delle popolazioni ittiche (Jennings *et al.* 2001).

L'incorporazione di queste informazioni 'qualitative' nello studio della biologia dei pesci e della pesca ha generato nuove intuizioni relative ai cambiamenti sia su scala temporale che spaziale delle popolazioni e comunità di pesci (Conover *et al.* 2006, Larsen *et al.* 2007). Queste nuove conoscenze hanno provocato un considerevole cambiamento nella nostra percezione dell'ambiente marino e di alcuni suoi abitanti chiave.

Gli studi genetici hanno consentito di dimostrare che la suddivisione di popolazione nei pesci marini è presente anche su scala geografica limitata, rivoluzionando la concezione classica di ambiente marino caratterizzato da una demografia 'aperta', in cui le specie possiedono elevata mobilità o potenziale di dispersione specialmente allo stadio di uova o larve. Il significato adattativo della struttura di popolazione nei pesci marini sta diventando ancora più evidente grazie all'individuazione di casi di divergenza nei tratti fenotipici, spesso anche in assenza di differenziamento genetico ai loci neutrali (Hutchings *et al.* 2007). Inoltre i dati pubblicati

negli ultimi anni indicano che le dimensioni effettive di popolazione (N_e) nei pesci marini, specialmente quelli caratterizzati da elevata fecondità e mortalità larvale, sono dai 2 ai 5 ordini di grandezza più piccole della popolazione di censo. Questo può significare che il numero di individui che si riproducono è molto più basso di quanto sarebbe immaginabile intuitivamente, e dunque che le popolazioni sono potenzialmente sensibili anche ad eventi demografici di portata ridotta. Le interessanti intuizioni fornite dai dati genetici hanno profonde implicazioni sia sulla comprensione della demografia delle specie marine, sia sulla loro gestione e conservazione.

Questo studio applica i metodi genetici allo studio del *Thunnus thynnus* e permette di consolidare aspetti della demografia su cui sono state formulate teorie già negli anni passati (Fromentin and Powers 2005), ma che dal punto di vista genetico fornivano risultati ambigui o comunque non accompagnati da evidenze chiare. Le analisi effettuate sui tonni provenienti dal bacino del Mediterraneo mostrano chiaramente la presenza di struttura tra i campioni con valori di un ordine di grandezza maggiore rispetto a studi precedenti (Carlsson *et al.* 2004, Carlsson *et al.* 2007). Inoltre i valori di F_{ST} globali nei campioni contemporanei e storici risultano quasi identici e suggeriscono che questo segnale di struttura nel Mediterraneo sia una caratteristica stabile nel tempo, fornendo ulteriori indizi alla teoria della metapopolazione (Fromentin and Powers 2005). Secondo questa teoria il tonno rosso può essere rappresentato come una collezione di popolazioni locali discrete e indipendenti che occupano habitat distinti ed irregolari e che presentano dinamiche proprie ma si influenzano tra loro attraverso fenomeni di dispersione. Questo segnale di struttura non è tuttavia abbastanza forte da essere individuato dal programma STRUCTURE alle condizioni sperimentali utilizzate (grandezza del campione e numero di loci). Alcuni studi hanno infatti dimostrato che con valori di F_{ST} uguali o inferiori a 0.03 il programma non riesce in modo efficiente ad individuare struttura (Latch *et al.* 2006). E' difficile valutare l'effetto del numero di loci considerato e della dimensione dei campioni, dunque su questo aspetto del nostro studio il giudizio va sospeso. Tuttavia è giustificato il sospetto che forse con un numero più elevato di loci il metodo sarebbe in grado di distinguere le sottopopolazioni. Il deficit di eterozigoti individuato nei campioni può essere interpretato come una forma di effetto Wahlund per cui nei campioni analizzati possono essere presenti individui appartenenti a unità riproduttive diverse. In ogni caso i livelli di differenziamento genetico osservato tra i campioni di *Thunnus thynnus* nel Mediterraneo centro-ovest sono più elevati rispetto anche a studi effettuati su scala geografica molto più ampia (Carlsson *et al.* 2007) e anche in altre specie di tonno (*Thunnus obesus*)

(Gonzalez *et al.* 2008). In base a questi risultati è possibile rifiutare con un discreto livello di confidenza l'ipotesi di panmissia tra i campioni studiati.

Dal punto di vista della conservazione delle specie la presenza di struttura all'interno di una popolazione può rappresentare un fattore di rischio di perdita di diversità se la specie in esame è fortemente sfruttata. Le sottopopolazioni che subiscono una riduzione in dimensione riescono a controbilanciare attraverso il flusso genico la maggiore perdita di variabilità dovuto all'effetto della deriva genetica. Se il flusso genico è ridotto, come suggerito dalla presenza di struttura, l'effetto della deriva risulta potenziato. In ogni caso le conseguenze genetiche a lungo termine sono peggiori per le metapopolazioni in quanto subiscono frequenti colli di bottiglia associati ad eventi regolari di estinzione e ricolonizzazione. Come già ampiamente illustrato nella introduzione la popolazione del tonno rosso è fortemente sfruttata da alcuni decenni. Perciò, per stabilire se la popolazione del Mediterraneo sia stata ridotta a livelli critici per il mantenimento della diversità genetica, sono state effettuate stime della dimensione della popolazione effettiva e della presenza di colli di bottiglia.

Confrontando il livello di eterozigosità, il numero di alleli e l'allelic richness nei campioni storici e contemporanei non si identifica una diminuzione chiara di variabilità in quelli attuali e i valori sono omogenei tra i diversi campioni. Anche la stima temporale della popolazione effettiva ($N_e \sim 1000$) indica che il tonno rosso possiede ancora un discreto potenziale genetico e dall'analisi a lungo termine emerge un quadro di stabilità della popolazione con valori di θ molto simili tra i campioni storici e contemporanei. Lo stesso quadro di stabilità emerge anche dalla stima del tasso di crescita che per tutti i campioni si avvicina a zero.

Dai valori di θ long-term stimati si sono ottenuti valori long-term di N_e molto più elevati rispetto a quelle temporali, con differenze di un ordine di grandezza anche usando un tasso di mutazione elevato (10^{-3}). Poiché la long-term N_e descrive la variazione genetica di una popolazione con una precisa demografia lungo la sua storia evolutiva, le stime long-term forniscono informazioni importanti sulla demografia storica del tonno rosso molto utili dal punto di vista della gestione della pesca. I valori di N_e inferiti con i due metodi sembrano riflettere rapporti N_e/N tipici dei pesci marini con una curva di sopravvivenza di tipo III, ossia alta fecondità, elevata mortalità giovanile ed elevata varianza del successo riproduttivo che può provocare un abbassamento di N_e . In effetti i valori di popolazione effettiva stimati sono molto bassi se si considerano le tonnellate di tonno pescato ogni anno che descrivono al contrario una popolazione molto ampia. Numerosi studi effettuati sui pesci marini indicano che il rapporto N_e/N medio è $\sim 10^{-4}$ (con un range 10^{-2} - 10^{-8} ; (Hauser and Carvalho 2008) e

che la dimensione effettiva di popolazione è da due a cinque ordini di grandezza più piccola della dimensione censo. Inoltre la relazione tra popolazione effettiva e popolazione di censo non può essere spiegata da una funzione lineare ma da una funzione esponenziale (Hauser and Carvalho 2008), questo implica che il rapporto N_e/N diminuisce con l'aumentare della popolazione e che anche in una popolazione di grandi dimensioni il numero di individui che si riproducono è bassa e perciò la varianza riproduttiva aumenta. Considerando la stima temporale ottenuta per i campioni di tonno del Mediterraneo centrale ($N_e = \sim 1000$) e il valore medio N_e/N (10^{-4}), la dimensione di censo della popolazione sarebbe intorno a 10^6 individui; utilizzando invece la stima long-term il valore si aggira intorno a 10^7-10^8 . Queste stime descrivono una popolazione molto ampia, dove tuttavia il numero di individui che si riproducono e danno origine ad una progenie vitale in proporzione è basso.

Dalle analisi della dimensione della popolazione effettiva non emerge quindi una situazione critica per la specie ($N_e < 500$ (Franklin and Frankham 1998)) che abbia intaccato il suo potenziale evolutivo nonostante il forte sfruttamento subito negli anni. Tuttavia va sottolineato che nell'effettuare queste stime non si è tenuto conto del tasso di migrazione che, specialmente in ambienti marini può avere una influenza sostanziale sia nella stima di N_e sia nel mantenimento di una elevata variabilità genetica, soprattutto quando le sottopopolazioni in esame presentano dei livelli bassi di struttura. Per valutare quindi se le stime ottenute di N_e possano non rappresentare la reale situazione di questa specie, dato che le assunzioni fatte (assenza di migrazione) possono aver prodotto sia una sottostima che una sovrastima della popolazione effettiva, sono stati effettuati dei test che consentono di stabilire se il grado di variabilità individuato nei campioni possa in realtà rappresentare un segnale di collo di bottiglia. I due test impiegati forniscono risultati molto contrastanti. Il metodo che calcola la presenza di eccesso/deficit di eterozigotità individua un segnale di bottleneck (eccesso di eterozigotità) in tutti i campioni solamente quando nelle analisi viene incorporato il modello mutazionale IAM che però non descrive in modo appropriato il processo di mutazione caratteristico dei microsatelliti. Quando invece vengono utilizzati i due modelli mutazionali che meglio si adattano ai microsatelliti (SSM, TPM), lo scenario cambia radicalmente e il test identifica dei segnali di espansione (deficit di eterozigotità) anche se non in tutti i campioni i risultati sono statisticamente significativi. Il metodo basato sul calcolo del rapporto M (numero di alleli/ range della dimensione degli alleli) fornisce dei risultati ambigui in quanto influenzati dal valore del parametro θ . Usando valori bassi come quelli ottenuti dalle stime temporali il metodo individua segnali di bottleneck in tutti i campioni (anche uno storico) tranne nelle due repliche del campione Adriatico. Inoltre alcuni dei valori di M sono di pochi

punti inferiori agli M critici stimati descrivendo quindi una situazione borderline e non danno un chiaro segnale di collo di bottiglia. Utilizzando i valori long-term di θ inoltre i valori di M stimati non sono mai significativi o inferiori ai rispettivi M critici. In generale non è possibile perciò indicare la presenza di un chiaro segnale di bottleneck con i due metodi utilizzati. Tuttavia considerando anche l'eterogeneità dei pattern genetici spaziali nel Mediterraneo queste evidenze contraddittorie tra i vari campioni rappresentano un ulteriore sostegno alla teoria che il bacino del Mediterraneo sia abitato da popolazioni in qualche modo distinte geograficamente, che mostrano una loro peculiare storia demografica.

La propensione all'indipendenza demografica nei tonni del Mediterraneo trova sostegno anche nella presenza di aree multiple adatte alla riproduzione nel Mediterraneo centrale (intorno alle isole Baleari e nel canale di Sicilia) e dai movimenti limitati mostrati dalla maggior parte dei tonni monitorati (Rooker *et al.* 2007, Rooker *et al.* 2008).

I risultati mostrati nelle tabelle 2 e 5 suggeriscono perciò che i campioni considerati hanno affrontato storie demografiche differenti.

Tuttavia non bisogna dimenticare che i metodi creati per individuare i colli di bottiglia spesso sono più adatti a riconoscere crisi demografiche di grossa entità, perciò il fatto che non identifichino un segnale chiaro a livello genetico potrebbe nascondere un evento importante dal punto di vista della conservazione di questo grosso pelagico. Riuscire a stabilire l'entità minima di un collo di bottiglia che può essere identificato da questi metodi significa riuscire a stimare quanto la dimensione della popolazione possa diminuire senza influenzare il numero di alleli in modo significativo. Chiarire questo punto resta un' importante priorità di ricerca, che verrà affrontata nel prosieguo del progetto di cui questa tesi rappresenta il primo stadio. Inoltre se consideriamo il fatto che in molti pesci marini la N_e è diversi ordini di grandezza inferiore alla popolazione totale, un evento di bottleneck potrebbe impoverire il gene pool provocando la perdita di potenziale evolutivo e quindi la capacità di adattamento ai cambiamenti ambientali. E' necessario perciò valutare con cautela i risultati ottenuti, anche in considerazione del fatto che gli alti livelli di diversità genetica potrebbero essere mantenuti attraverso il flusso genico generato dalla migrazione che può controbilanciare l'effetto della deriva genetica.

6. BIBLIOGRAFIA

- Alvarado Bremer JR, Naseri I, Ely B. (1999) Heterogeneity of northern bluefin tuna populations. *ICCAT Col Vol Sci Pap*; **49**: 127–129.
- Alvarado Bremer JR, Vinas J, Mejuto J, Ely B, Pla C. (2005) Comparative phylogeography of Atlantic bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. *Mol Phylogenet Evol*; **36**: 169-187.
- Anderson EC, Williamson EG, Thompson EA. (2000) Monte Carlo Evaluation of the Likelihood for Ne From Temporally Spaced Samples *Genetics*; **156**: 2109-2118.
- Beaumont MA, Rannala B. (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics*; **5**: 251-261.
- Berli P, Felsenstein J. (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc Natl Acad Sci U S A*; **98**: 4563-4568.
- Berthier P, Beaumont MA, Cornuet JM, Luikart G. (2002) Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics*; **160**: 741-751.
- Block BA, Dewar H, Blackwell SB, Williams TD, Prince ED, Farwell CJ *et al.* (2001a) Migratory movements, depth preferences, and thermal biology of Atlantic bluefin tuna. *Science*; **293**: 1310–1314.
- Block BA, Teo SL, Walli A, Boustany A, Stokesbury MJ, Farwell CJ *et al.* (2005) Electronic tagging and population structure of Atlantic bluefin tuna. *Nature*; **434**: 1121-1127.
- Boustany AM, Reeb CA, Teo SL, De Metrio G, Block BA. (2006) Genetic data and electronic tagging indicate that the Gulf of Mexico and Mediterranean Sea are reproductively isolated stocks of bluefin tuna (*Thunnus thynnus*). *SCRS* **089**:
- Carlsson J, McDowell JR, Carlsson JE, Graves JE. (2007) Genetic identity of YOY bluefin tuna from the eastern and western Atlantic spawning areas. *J Hered*; **98**: 23-28.
- Carlsson J, McDowell JR, Diaz-Jaimes P, Carlsson JE, Boles SB, Gold JR *et al.* (2004) Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) population structure in the Mediterranean Sea. *Mol Ecol*; **13**: 3345-3356.
- Conover DO, Clarke LM, Munch SB, Wagner GN. (2006) Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *Journal of Fish Biology*; **69**: 21–47.
- Cornuet JM, Luikart G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*; **144**: 2001-2014.
- Corriero A, Karakulak S, Santamaria N, Deflorio M, Spedicato D, Addis P *et al.* (2005) Size and age at sexual maturity of female bluefin tuna (*Thunnus thynnus* L. 1758) from the Mediterranean Sea. *Journal of Applied Ichthyology*; **21**: 483 - 486.
- Cowen RK, Paris CB, Srinivasan A. (2006) Scaling of connectivity in marine populations. *Science*; **311**: 522-527.
- De Metrio G, Arnold GP, Arnold GP, Block BA, Megalafonou P, Lutcavage M *et al.* (2005) Movements of bluefin tuna (*Thunnus thynnus* L.) tagged in the Mediterranean Sea with pop-up satellite tags. . *ICCAT Col Vol Sci Pap*; **58**: 1337–1340.
- El Mousadik A, Petit RJ. (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. . *Theor Appl Genet* **92**: 832-839.
- Excoffier L. (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**: 47-50.

- Excoffier L, Smouse P, Quattro J. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*; **131**: 479-491.
- Falush D, Stephens M, Pritchard JK. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*; **164**: 1567-1587.
- Frankham R. (2005) Stress and adaptation in conservation genetics. *J Evol Biol*; **18**: 750-755.
- Franklin IR, Frankham R. (1998) How large must populations be to retain evolutionary potential? *Animal Conservation*; **1**: 69-73.
- Fraser DJ, Hansen MM, Ostergaard S, Tessier N, Legault M, Bernatchez L. (2007) Comparative estimation of effective population sizes and temporal gene flow in two contrasting population systems. *Mol Ecol*; **16**: 3866-3889.
- Fromentin JM, Powers JE. (2005) Atlantic bluefin tuna: population dynamics, ecology, fisheries and management. *Fish Fisheries* **6**: 281-306
- Garcia A, Alemany F, Velez-Belchi P, Rodriguez JM, Lopez Jurado JL, Gonzalez Pola C *et al.* (2003) Bluefin and frigate tuna spawning off the Balearic archipelago in the environmental conditions observed during the 2002 spawning season. *ICCAT Col Vol Sci Pap*; **55**: 1261-1270.
- Garza JC, Williamson EG. (2001) Detection of reduction in population size using data from microsatellite loci. *Mol Ecol*; **10**: 305-318.
- Gonzalez EG, Beerli P, Zardoya R. (2008) Genetic structuring and migration patterns of Atlantic bigeye tuna, *Thunnus obesus* (Lowe, 1839). *BMC Evol Biol*; **8**: 252.
- Goudet J. (1995) Fstat version 1.2. A computer program to calculate F-statistics. *J Hered*; **86**:
- Griffiths RC, Tavaré S. (1994) Simulating probability distributions: estimation using the likelihood of changes in marker allele distributions in the coalescent. *Theor Popul Biol*; **46**: 131-159.
- Guinand B, Scribner KT. (2003) Evaluation of methodology for detection of genetic bottlenecks: inferences from temporally replicated lake trout populations. *C R Biol*; **326 Suppl 1**: S61-67.
- Hastings WK. (1970) Monte Carlo sampling methods using Markov chains and their application. *Biometrika*; **57**: 97-109.
- Hauser L, Adcock GJ, Smith PJ, Ramirez JH, Carvalho GR. (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proc Natl Acad Sci U S A*; **99**: 11742-11747.
- Hauser L, Carvalho GR. (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fish*; **9**: 333-362.
- Hudson RR. (1990) Genetic Data Analysis. Methods for Discrete Population Genetic Data. Bruce S. Weir. Sinauer, Sunderland, MA, 1990. xiv, 377 pp., illus. \$48; paper, \$27. *Science*; **250**: 575.
- Hutchings JA, Swain DP, Rowe S, Eddington JD, Puvanendran V, Brown JA. (2007) Genetic variation in life-history reaction norms in a marine fish. *Proceedings of the Royal Society of London Series B: Biological Sciences*; **274**: 1693-1699.
- ICCAT. (2007) Report of the 2006 Atlantic bluefin tuna stock assessment session (SCRS/2006/013). *International Commission for the Conservation of Atlantic Tuna (ICCAT)*; **Madrid** 137 pp.
- Jennings S, Kaiser MJ, Reynolds JD. (2001) Marine Fisheries Ecology. *Blackwell Publishing, Malden, MA*:
- Kimura M, Crow JF. (1964) The Number of Alleles That Can Be Maintained in a Finite Population. *Genetics*; **49**: 725-738.
- Kingman JFC. (1982a) On the Genealogy of Large Populations *Journal of Applied Probability*; **19**: 27-43.

- Kingman JFC. (1982b) The coalescent. *Stochastic Processes and Their Applications* **13**: 235-248.
- Kingman JFC. (1982c) Exchangeability and the evolution of large populations. *Exchangeability in Probability and Statistics Proceedings of the International Conference on Exchangeability in Probability and Statistics, Rome, 6th-9th April, 1981*: 97-112
- Kuhner MK. (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics*; **22**: 768-770.
- Kuhner MK, Yamato J, Felsenstein J. (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*; **149**: 429-434.
- Larsen PF, Nielsen EE, Williams TD, Hemmer-Hansen J, Chipman JK, KRUHØFFER M *et al.* (2007) Adaptive differences in gene expression in European flounder (*Platichthys flesus*). *Molecular Ecology*; **16**: 4674-4683.
- Latch E, Dharmarajan G, Glaubitz JC, Rhodes OE J. (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*; **7**: 295-302.
- Marjoram P, Tavaré S. (2006) Modern computational approaches for analysing molecular genetic variation data. *Nat Rev Genet*; **7**: 759-770.
- Marshall CT, Browman HI. (2007) Disentangling the causes of maturation trends in exploited fish populations. *Marine Ecology Progress Series*; **335**: 249-251.
- Metropolis N, Rosenbluth AN, Rosenbluth MN, H. TA, Teller E. (1953) Equations of state calculations by fast computing machine. *J Chem Phys*; **21**: 1087-1091.
- Nishida T, Tsuji S, Segawa K. (1998) Spatial data analyses of Atlantic bluefin tuna larval surveys in the 1994 ICCAT BYP. *ICCAT Col Vol Sci Pap*; **48**: 107-110.
- Ohta T, Kimura M. (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. *Genet Res* **22**: 201-204.
- Oray IK, Karakulak SF. (2005) Further evidence of spawning of bluefin tuna (*Thunnus thynnus* L., 1758) and the tuna species (*Auxis rochei* Ris., 1810, *Euthynnus alletteratus* Raf., 1810) in the eastern Mediterranean Sea: preliminary results of TUNALEV larval survey in 2004. *Journal of Applied Ichthyology*; **21**: 236-240.
- Petit RJ, El Mousadik A, Pons O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*; **12**: 844-855.
- Piry S, Luikart G, Cornuet JM. (1999) BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *The Journal of Heredity*; **90**: 502-503.
- Pritchard JK, Stephens M, Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics*; **155**: 945-959.
- Rambaut A, Drummond AJ. (2004) TRACER. Version 1.3. <http://beastbioedacuk/Tracer>:
- Raymond M. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered*; **86**: 248-249.
- Rice WR. (1989) Analyzing Tables of Statistical Tests. *Evolution*; **43**: 223-225
- Ripley BD. (1987) Stochastic Simulation. *John Wiley & Sons*: New York.
- Rooker JR, Alvarado Bremer JR, Block BA, Dewar H, De Metrio G, Corriero A *et al.* (2007) Life history and stock structure of Atlantic Bluefin Tuna (*Thunnus thynnus*). *Reviews in Fisheries Science*; **15**: 265-310.
- Rooker JR, Secor DH, De Metrio G, Schloesser R, Block BA, Neilson JD. (2008) Natal homing and connectivity in Atlantic bluefin tuna populations. *Science*; **322**: 742-744.
- Rosenberg NA, Nordborg M. (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*; **3**: 380-390.
- Ross SM. (1997) Simulation. *Academic Press, San Diego*: Second Edition.

- Ruzzante DE, Mariani S, Bekkevold D, Andre C, Mosegaard H, Clausen LA *et al.* (2006) Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. *Proc Biol Sci*; **273**: 1459-1464.
- Sella M. (1927) Les migrations des thons etudiees par le moyen des hamecons. *Bull Sta Aquic Pech Castiglione*; **2**: 101–136.
- Sella M. (1929b) Migrazione e habitat del tonno (*Thunnus thynnus* L.) studiati col metodo degli ami, con osservazioni sull'accrescimento, sul regime delle tonnare ecc. *Mem R Com Talassogr Ital*; **156**: 1–24.
- Spong G, Hellborg L. (2002) A near-extinction event in lynx: do microsatellite data tell the tale? . *Conservation Ecology* **6**: 15.
- Teo SLH, Boustany A, Dewar H, Stokesbury MJW, Weng KC, Beemer S *et al.* (2007) Annual migrations, diving behavior, and thermal biology of Atlantic bluefin tuna, *Thunnus thynnus*, on their Gulf of Mexico breeding grounds. *Marine Biology* **151**: 1-18.
- Treml EA, Halpin PN, Urban DL, Pratson LF. (2008) Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. *Landscape Ecology*; **23**: 19–36.
- Waldick RC, Kraus S, Brown M, White BN. (2002) Evaluating the effects of historic bottleneck events: an assessment of microsatellite variability in the endangered, North Atlantic right whale. *Mol Ecol*; **11**: 2241-2249.
- Wang J. (2001) A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genet Res, Camb*; **78**: 243-257.
- Waples RS, Adams PB, Bohnsack J, Taylor BL. (2007) A biological framework for evaluating whether a species is threatened or endangered in a significant portion of its range. *Conserv Biol*; **21**: 964-974.
- Waples RS, Yokota M. (2007) Temporal estimates of effective population size in species with overlapping generations. *Genetics*; **175**: 219-233.
- Wright S. (1938) Size of population and breeding structure in relation to evolution. *Science*; **87**: 430-431.
- Wright S. (1969) Evolution and the Genetics of Populations. *University of Chicago Press, Chicago*; **Vol. 2.**: