



Università degli Studi di Firenze

DOTTORATO DI RICERCA IN
"Scienze genetiche, microbiologiche e
bioinformatica"

CICLO XXV

COORDINATORE Prof.ssa Milvia Luisa Racchi

Analisi della diversità genetica in due specie
arboree mediterranee: palma da dattero
(*Phoenix dactylifera* L.) e pino nero (*Pinus nigra*
Arnold)

Settore Scientifico Disciplinare AGR07/BIO18

Dottorando

Dott. Bove Andrea

Tutore

Prof. Racchi Milvia L.

Cotutore

Dott. Vendramin Giovanni G.

Anni 2010/2012

Dichiarazione

Con la presente affermo che questa tesi è frutto del mio lavoro e che, per quanto io ne sia a conoscenza, non contiene materiale precedentemente pubblicato o scritto da un'altra persona né materiale che è stato utilizzato per l'ottenimento di qualunque altro titolo o diploma dell'Università o altro istituto di apprendimento, a eccezione del caso in cui ciò venga riconosciuto nel testo.

Una copia della tesi sarà disponibile presso DiBA (DISPAA dal 1° gennaio).

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by other person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

A copy of the thesis will be available at DiBA (DISPAA from January 1st 2013)

Firenze, 31 dicembre 2012

Andrea Bove

Riassunto

Questo progetto di ricerca del dottorato si articola su due casi di studio simili nelle metodiche per generare i dati, basate su tecniche di genotipizzazione molecolare, ma con approcci differenti nell'analisi e soprattutto nell'interpretazione dei dati. In entrambi i casi di studio la diversità genetica è stata stimata mediante marcatori molecolari (microsatelliti-SSR-*short sequence repeat*- e SNP-*single nucleotide polymorphism*), ma mentre nel caso della palma da dattero è utilizzata per caratterizzare il germoplasma, nel caso del pino nero costituisce il prerequisito per la comprensione del ruolo relativo dei fenomeni selettivi e demografici nel definire la struttura genetica delle popolazioni, per studi di associazione molecolare, di filogeografia e più in generale di genetica della conservazione.

Il primo caso di studio riguarda le indagini avviate in Libia sulla caratterizzazione della diversità genetica della palma da dattero (*Phoenix dactylifera*). Il germoplasma fenicolo libico preserva ancora oggi un'enorme ricchezza comprovata dalle oltre 400 varietà esistenti e descritte, fino ad ora, solamente utilizzando i caratteri morfologici come chiave identificativa delle cultivar. Abbiamo analizzato 377 piante distribuite principalmente in 3 località della regione di Al-Jufra, appartenenti a 18 cultivar, la maggior parte di derivazione autoctona. E' stato inserito nelle analisi anche un sottogruppo di piante maschili (n=63) di varietà ignota che rappresentano i principali impollinatori dell'oasi. La variabilità genetica è stata valutata utilizzando un set di 16 marcatori SSR che mappano loci nucleari e un locus minisatellite plastidiale. Abbiamo registrato un elevato numero di alleli SSR, con una media di 6,88 alleli per locus che hanno permesso di rivelare un alto grado di variabilità genetica in queste accessioni. Questi alti livelli di polimorfismo presenti nelle cultivar sono in accordo con quelli precedentemente riportati per le cultivar distribuite nelle regioni Nord Africane e Medio Orientali. E' stata costruita una chiave identificativa delle cultivar utilizzando tre loci SSR altamente polimorfici, che hanno esibito 23 alleli in totale. Tutte le cultivar sono state identificate in maniera semplice e non ambigua sulla base dei loro profili allelici. L'analisi della struttura genetica ha messo in evidenza che 9 cultivar su 18 hanno un'identità clonale del 100% tra le diverse località dell'oasi evidenziando l'ottima abilità, maturata grazie all'esperienza tramandata negli anni dagli agricoltori, nelle pratiche di propagazione

clonale delle cultivar; tuttavia abbiamo registrato nel nostro campione due casi di classificazione errata, che si possono verificare durante la propagazione a causa della difficoltà nella distinzione di alcune cultivar, soprattutto quelle strettamente correlate, sulla base dei descrittori pomologici. I risultati delle analisi effettuate sul locus minisatellite plastidiale hanno evidenziato che le cultivar autoctone libiche presentano tutte l'aplotipo a tre ripetizioni. Analisi precedenti, hanno mostrato che la specie dactylifera presenta due aplotipi del minisatellite, a tre o quattro ripetizioni, che potrebbe rispecchiare l'origine, occidentale o orientale rispettivamente, del materiale fenicicolo; e quindi un'ipotetica origine occidentale delle palme libiche. Nella prospettiva futura di migliorare la qualità del frutto attraverso un piano di incrocio ricorrente, le piante maschili hanno bisogno di essere identificate e assegnate a una specifica cultivar. Di conseguenza, la possibilità di attribuire a una pianta maschile di varietà sconosciuta la probabile varietà di appartenenza è stato effettuato attraverso analisi di parentela. I risultati ottenuti permettono di assegnare con un elevato livello di confidenza 24 maschi sui 63 analizzati, a una delle 18 cultivar. I risultati positivi ottenuti nell'identificazione delle piante maschili confermano ulteriormente l'utilità dei microsattelliti per le analisi di genotipizzazione e aprono nuove prospettive per il miglioramento in questa specie.

L'efficacia e la potenza discriminatoria mostrata dai microsattelliti nell'individuazione di tutte le varietà e accessioni esaminate, conferma l'utilità di questi marcatori nelle tecniche di "fingerprinting" clonale e identificazione varietale. Il trasferimento di questi marcatori può risultare di grande utilità per l'etichettatura su larga scala dei polloni e delle colture in vitro. I nostri dati evidenziano la possibilità di usare questi potenti marcatori come descrittori nella certificazione e nel controllo di origine del materiale vegetale.

Lo studio sul pino nero ha richiesto l'identificazione di un set di geni candidati e l'analisi di SNP (*single nucleotide polymorphism*). La valutazione della distribuzione della diversità genetica in popolazioni naturali è di fondamentale importanza; infatti, la variabilità genetica permette agli organismi vegetali di sopravvivere ed adattarsi ai cambiamenti ambientali, in particolare quando sono repentini, come di fatto sta accadendo ora. Per questa ragione la comprensione del ruolo dei geni responsabili dei

meccanismi legati all'adattamento assume grande importanza. Analoga rilevanza riveste l'analisi dei polimorfismi a livello di questi geni in popolamenti naturali campionati lungo gradienti geografici, e la valutazione del ruolo relativo della demografia (espansione e contrazione della popolazione) e della selezione sulla distribuzione di questi polimorfismi. In questo lavoro, che rientra in un'iniziativa internazionale di risequenziamento di geni candidati delle principali conifere europee, è stata valutata la diversità genetica in 7 popolazioni costituite da 12 individui ciascuna di *Pinus nigra*, una specie montana mediterranea. La variabilità nucleotidica è stata stimata tramite marcatori molecolari SNP a livello di 14 geni, selezionati da un dataset di 264 geni candidati identificati su *P. taeda*, una specie di pino Nord-Americano. A livello delle sequenze geniche analizzate sono state identificate sia mutazioni puntiformi (SNP) che inserzioni-delezioni, per un totale di 270 polimorfismi. La diversità nucleotidica calcolata come valore medio sui 14 geni analizzati è risultata pari a 0,0058, in accordo con i valori medi ritrovati in altre specie del genere *Pinus*. Le popolazioni appartenenti all'areale orientale di distribuzione della specie manifestano una maggiore diversità rispetto alle popolazioni occidentali. I test di neutralità hanno evidenziato la presenza di geni che discostano dal modello neutrale e che manifestano segnali di selezione. Metodi di classificazione Bayesiana hanno messo in evidenza la presenza di clusters che possono essere ricondotti alla storia di questa specie, in particolare a quella che fa riferimento all'ultima glaciazione. A livello tassonomico è ancora molto dibattuta la questione della possibile presenza di sottospecie. In questo lavoro abbiamo verificato le possibili affinità tra le popolazioni campionate, confrontando popolazioni che ricoprono la maggior parte dell'areale di distribuzione Mediterraneo. La distribuzione della diversità nucleotidica sembra in parte essere riconducibile alla presenza di sotto-specie e confermare lo status di complex del *Pinus nigra*. La strategia di SNP-discovery in geni candidati, utilizzata in questo lavoro, è al momento la migliore in specie appartenenti alla famiglia delle Pinaceae dove la mancanza di un genoma di riferimento e le dimensioni enormi dei genomi stessi non consentono la scelta di altre strategie (ad esempio il re-sequenziamento di interi genomi) per monitorare la diversità genetica su ampia scala (wide genome scan).

Abstract

The Ph.D. thesis refers to two case studies where similar methods to generate the molecular data but different approaches of data analyses and interpretation were adopted. In both cases the genetic diversity has been estimated by molecular markers (microsatellites-SSR-short sequence repeat and SNP-single nucleotide polymorphism). In the date palm case the genetic variability is used for the characterization of germplasm, while in the *Pinus nigra* case a more detailed analysis for understanding the relative role of selection and demography on the genetic structure of populations, for molecular association and phylogeographic studies were performed.

The first part of the thesis concerns the characterization of the genetic diversity of date palm (*Phoenix dactylifera*) in Libya. Date palm germplasm in Libya still preserves an enormous richness, attested by the presence of more than 400 different accessions, widely described on the basis of morphological characters. 377 plants mainly sampled in three locality of Al-Jufrah oasis, belonging to 18 cultivars were analysed. An analysis of male trees (n=63) of unknown cultivar, representing the best pollinators in this region, was also included. Genetic variability was evaluated using a set of 16 highly polymorphic nuclear microsatellite loci and one plastidial minisatellite locus. A high number of SSR alleles were detected with a mean of 6.88 per locus. A high level of differentiation among cultivars was detected, as previously reported in other studies. An identification key was built using three microsatellite loci that revealed 23 alleles in total. All local cultivars were unambiguously identified on the basis of the allelic profiles. 9 out of 18 cultivars display no genetic diversity within the cultivar, as a result of clonal propagation of the cultivars. Cases of misclassification can occur during propagation because of the difficulty to identify some cultivars on the basis of morphological traits.

The minisatellite locus showed that all the autochthonous Libyan date palm share the haplotype displaying three repeats. Previous analysis showed that date palm presents two minisatellites haplotypes, with three or four repeats, that may reflect the origins, western vs eastern respective, of the vegetal materials; and then a hypothetical western origin of the Libyan date palm.

To improve the fruit quality through a recurrent breeding procedure, the male tree pollinators need to be genetically identified and assigned to a specific

cultivar. Consequently, the possibility to attribute the unknown male tree to a cultivar was considered and parentage analysis was performed on male trees. The results obtained allowed to assign with high confidence level 24 male trees out of 63 to a specific cultivar.

The effectiveness of SSR in discriminating all the examined accessions and cultivars confirm the usefulness of these markers for clone fingerprinting and cultivar identification. Since each variety was identified by a unique profile, it is possible to generate an individual barcode using the multi-locus genotype for the certification and the control of origin labels of date palm products.

The second case study involves the SNP-discovery in a set of candidate genes. The evaluation of the distribution of genetic diversity in natural populations is of fundamental importance, since this genetic variability enables living organisms to survive in an environment where change is abrupt. It thus becomes important to understand the role of genes related to adaptation mechanisms and estimate polymorphism in populations existing along geographical gradients of the target species, in order to highlight whether these polymorphisms are due to demographic (expansion and contraction of population) or are the result of selective events. In this work, which is part of an international re-sequencing project of a large number of candidate genes on the main European conifers, we assessed the distribution of genetic diversity within and among natural populations of a mountain Mediterranean species, *Pinus nigra*. The nucleotide variability was assessed by SNP markers (Single Nucleotide Polymorphism) on 14 genes, selected from a set of 265 candidate genes identified on *P. taeda*, a Nord-American pine species.

Both point mutations (SNP) that insertions-deletions, for a total of 270 polymorphisms, were identified at the gene level. The nucleotide diversity calculated as the average value of the 14 genes analyzed was 0.0058, in agreement with the values found in other species of the genus *Pinus*. The populations of eastern part of the distribution of the species show an higher diversity than the western populations. Tests of neutrality showed the presence of genes that deviate from the neutral model and which manifest signals of selection. Bayesian classification methods revealed the presence of clusters that can be traced to the history of this species, in particular that referring to the last glaciation.

The possible presence of subspecies is a still debated issue. In this work, the possible relationships among the sampled populations were clarified. The distribution of nucleotide diversity seems to be partly related to the presence of sub-species and confirm the status of a complex species of *Pinus nigra*. The strategy of SNP-discovery in candidate genes, used in this work, is currently the best one in species belonging to the family of Pinaceae where the lack of a reference genome and the enormous size of the genomes do not allow the choice of other strategies (e.g. re-sequencing of entire genomes) to monitor genetic diversity on a large scale (genome wide scan).

INDICE

Introduzione generale	1
1. Valutazione della diversità genetica	1
1.1 Marcatori molecolari microsatelliti: caratteristiche ed utilizzo	5
1.2 Marcatori molecolari SNP: caratteristiche ed utilizzo	8
Capitolo 1	11
<i>Analisi della diversità genetica della palma da dattero (Phoenix dactylifera) nell'Oasi di Al-Jufrah (Libia)</i>	
1. Introduzione	11
1.1 Introduzione sul genere Phoenix: tassonomia e filogenetica	12
1.2 Phoenix dactylifera	15
1.3 Caratterizzazione varietale	21
1.4 Sequenziamento e caratterizzazione del genoma di <i>P.dactylifera</i>	23
1.5 Domesticazione della palma da dattero	26
1.6 La fienicoltura libica	28
2. Scopo della tesi	32
3. Materiali e Metodi	33
3.1 Descrizione del materiale campionato	33
3.2 Estrazione del DNA	36
3.3 Amplificazione e genotipizzazione dei loci SSRs nucleari	36
3.4 Genotipizzazione del locus minisatellite plastidiale	38
3.5 Analisi dei dati	39
4. Risultati	41

4.1 Analisi della diversità genetica	41
4.2 Chiave identificativa delle 18 cultivar	51
4.3 Assegnazione delle piante maschili ad una varietà	53
4.4 Analisi del minisatellite plastidiale	56
5. Discussione	60
Ringraziamenti	63
Capitolo 2	64
<i>SNP in geni candidati per valutare la diversità genetica in popolazioni naturali di Pinus nigra</i>	
1. Introduzione	64
1.1 Introduzione sul genere <i>Pinus</i> : tassonomia e filogeografia	65
1.2 <i>Pinus nigra</i>	77
2. Scopo della tesi	83
3. Materiali e Metodi	84
3.1 Generazione di un dataset di geni orologi candidati	84
3.1.1 Campione utilizzato	84
3.1.2 Geni candidati analizzati	85
3.1.3 Analisi dei dati	86
3.2 Ri-sequenziamento di 20 geni candidati in sette popolazioni Mediterranee di <i>Pinus nigra</i>	87
3.2.1 Campione utilizzato	87
3.2.2 Scelta e amplificazione dei geni candidati	89
3.2.3 Analisi dei dati	90

4. Risultati	94
4.1 Generazione di un dataset di geni ortologi candidati	94
4.2 Ri-sequenziamento di 20 geni candidati in sette popolazioni Mediterranee di <i>Pinus nigra</i>	96
4.2.1 Stima della diversità e differenziazione nucleotidica	98
4.2.2 Test di neutralità	103
4.2.3 Indagine molecolare a livello delle sottospecie	108
5. Discussioni	118
Ringraziamenti	123
Riferimenti bibliografici	124

INTRODUZIONE GENERALE

1. Valutazione della diversità genetica.

La diversità genetica è la variabilità presente nel genoma delle specie (variazioni di sequenza DNA) Essa costituisce il materiale su cui agiscono i fattori evolutivi e permette alle specie di evolvere ed adattarsi a nuove situazioni, come ad esempio quelle legate ai cambiamenti ambientali. E' stato ampiamente riconosciuto che la perdita di diversità genetica può condurre alla riduzione della fitness degli individui e delle popolazioni e rappresenta una delle principali minacce per il mantenimento del potenziale adattativo delle specie (Frankham 1997).

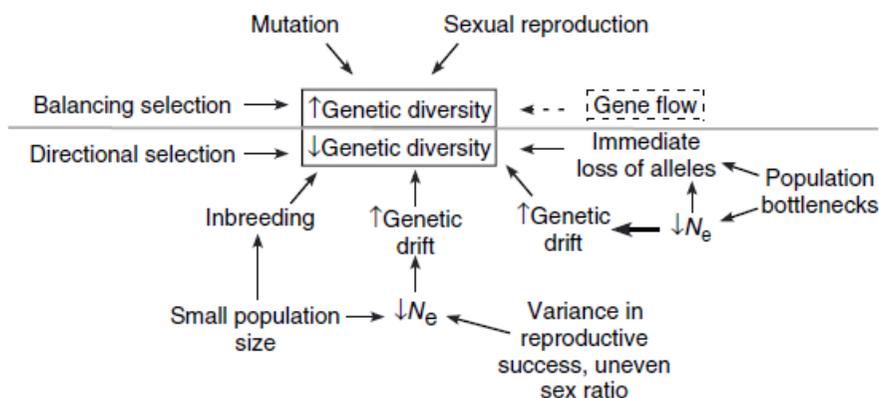
La diversità genetica tradizionalmente è stimata su due livelli di organizzazione, all'interno e tra le popolazioni della specie in esame, dove con popolazione si intende il più piccolo insieme di organismi di una singola specie capaci di scambiare geni attraverso la riproduzione sessuale nello stesso tempo e nello stesso spazio. La popolazione costituisce un'unità di riproduzione, la cui dimensione dipende dalla distanza geografica e dai modi di dispersione del polline e del seme.

La variazione della diversità genetica all'interno e tra le popolazioni naturali è il risultato di complessi processi di mutazione, selezione, deriva genica, migrazione, sistemi di riproduzione e dimensione effettiva della popolazione (N_e), vale a dire di tutti quei fattori demografici e selettivi. La figura 1 riassume i principali fattori che influenzano i livelli di diversità genetica all'interno e tra le popolazioni. Ad esempio il fenomeno del collo di bottiglia (bottleneck), dovuto ad una drastica riduzione della dimensione effettiva della popolazione provoca la perdita di alleli dal pool genico della specie o conduce alla fissazione di un allele. La deriva genetica casuale, che ha effetti maggiori su piccole popolazioni, può essere il risultato di eventi stocastici ambientali (periodi glaciali, incendi, inondazioni) ma anche fenomeni mediati dall'uomo (frammentazione degli habitat, domesticazione delle specie agrarie, riscaldamento globale), che portano ad una diminuzione della diversità genetica.

Le specie arboree sono importanti riserve di diversità genetica, individuale e di comunità, come conseguenza della loro complessa storia evolutiva e della variazione ambientale su scala locale e regionale. Diviene fondamentale

stimare la quantità di diversità genetica presente in queste specie sia per i programmi di miglioramento del germoplasma che per la messa a punto di strategie di conservazione genetica *in-situ* ed *ex-situ* con implicazioni nel comparto agrario e forestale.

Figura 1: Principali fattori che influenzano i livelli di diversità genetica all'interno delle popolazioni (Fonte: Freeland 2005).



Alcuni modelli previsionali (IPCC 2007a, b) evidenziano come gli impatti dei cambiamenti globali sulle specie arboree saranno acuti, in particolare nelle regioni Mediterranee (Giorgi & Lionello 2008), con notevoli cambiamenti nella distribuzione delle specie, nelle funzioni degli ecosistemi e nell'interazione tra i vari organismi. In aggiunta, se si considera che le regioni Mediterranee ospitano un'enorme ricchezza di biodiversità rispetto al continente, queste problematiche risultano ancora più accentuate (Fady 2005). Viste le caratteristiche sessili e di longevità, le specie arboree, prima di estinguersi, possono far fronte a questi cambiamenti attraverso la migrazione, mediata dai fenomeni dispersivi (via seme o polline), in ambienti dove le condizioni ecologiche sono più ospitali o adattarsi *in situ* in tempi rapidi sfruttando la variabilità genetica esistente. E' stato dimostrato che l'adattamento ai rapidi cambiamenti climatici, negli organismi arborei, risiede principalmente nella variabilità genetica esistente e nella ricombinazione piuttosto che su nuove mutazioni (Aitken et al., 2008). Questo non solo perché alleli favorevoli sono immediatamente disponibili, ma anche perché un allele favorevole presente nel genoma è più antico di

una mutazione recente e potrebbe essere già stato valutato dalla selezione in ambienti passati, in un'altra parte dell'areale di distribuzione della specie o anche in un'altra specie con la quale la popolazione ha scambiato geni (Aitken et al., 2008).

Un'efficiente gestione e conservazione della diversità genetica passa quindi inevitabilmente dalla possibilità di valutarla.

L'introduzione dei marcatori molecolari ha permesso un'accurata valutazione della diversità genetica, precedentemente stimata indirettamente a livello fenotipico. I marcatori molecolari possiedono numerosi vantaggi rispetto ai caratteri morfologici in quanto sono altamente ereditabili, non sono influenzati dagli stadi di sviluppo e dall'ambiente e sono relativamente facili da saggiare (Duran et al., 2009).

I primi strumenti molecolari utilizzati per valutare la diversità genetica in specie arboree hanno riguardato il monitoraggio di vari loci isoenzimatici (Torres & Tisserat 1980; Hamrick & Godt 1990; Morgante & Vendramin 1991), i quali hanno il vantaggio di essere codominanti e trasferibili, ma quelli altamente informativi sono poco abbondanti nel genoma. L'avvento dei marcatori basati sul DNA ha rivoluzionato il campo della genetica delle popolazioni perché il numero di loci e di polimorfismi rilevabili è pressoché illimitato.

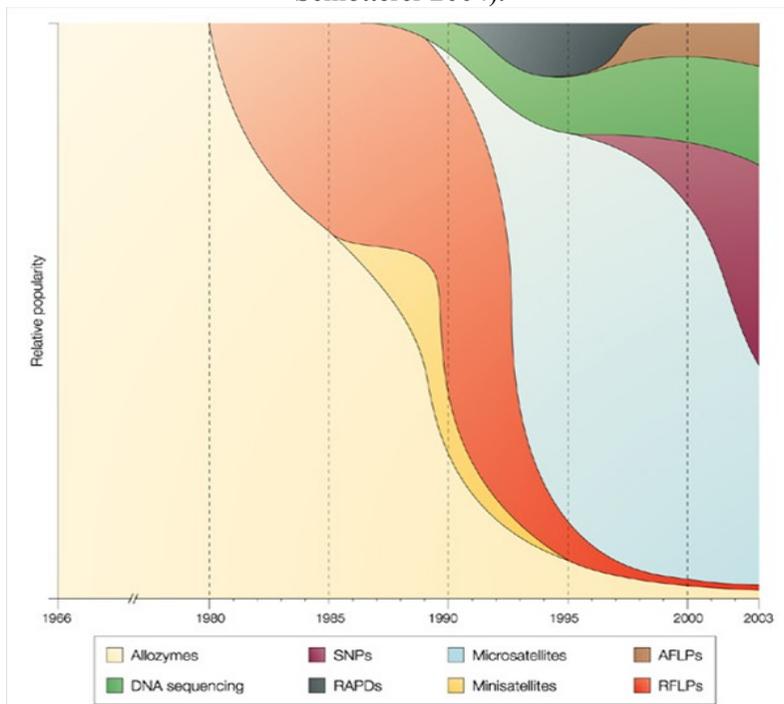
I marcatori molecolari rivelano i polimorfismi a livello di sequenza di DNA causati da mutazioni per inserzioni, delezioni, duplicazioni nucleotidiche. Questi marcatori possono presentare ereditarietà biparentale quando si ritrovano a livello del genoma nucleare e se codominanti si può chiaramente identificare la condizione omozigote o eterozigote dell'individuo. I marcatori molecolari hanno inoltre un'ereditarietà uniparentale che può essere materna, per la maggior parte dei genomi mitocondriali e plastidiali nelle Angiosperme, o paterna, per i genomi plastidiali della maggior parte delle conifere e in particolare per la famiglia delle Pinaceae (Neale & Sederoff 1989). Questa peculiarità dei genomi plastidiali può essere sfruttata per identificare marcatori seme-specifici e polline-specifici rispettivamente per le Angiosperme e le Gimnosperme (Powell et al., 1996).

Con la scoperta della reazione a catena della polimerasi (PCR) è comparsa tutta quella classe di marcatori PCR dipendenti che mettono in risalto sia i polimorfismi nelle sequenze del DNA (SNP, single nucleotide polymorphism) che le variazioni nelle ripetizioni del DNA (Schlotterer,

2004). Tra questi, sulla base della specificità dei primer, si distinguono marcatori multilocus come i RAPD (random amplified polymorphic DNA) ottenuti con un unico primer o gli AFLP (amplification fragment length polymorphism) ottenuti con una coppia di primer, e marcatori single-locus come gli SSR, (simple sequence repeat) che permettono di visualizzare il polimorfismo, in un gruppo di individui della medesima specie o genere, di una singola regione del genoma compresa tra una coppia di primers altamente specifici. I marcatori che mettono in risalto le variazioni nelle ripetizioni del DNA sono considerati selettivamente neutri dato che generalmente sono localizzati in regioni non codificanti. I marcatori SNPs invece possono essere localizzati in regioni codificanti o regolative nel qual caso possono subire pressioni selettive (Awadalla & Ritland 1997).

Le diverse classi di marcatori molecolari hanno trovato impiego e diffusione diversi nel corso dello sviluppo delle metodiche di ricerca (Figura 2).

Figura 2: Popolarità relativa delle principali classi di marcatori molecolari (Fonte: Schlotterer 2004).



Oggigiorno nelle analisi genetiche sono predominanti i marcatori SSRs e SNPs, che costituiscono le tipologie di marcatori molecolari utilizzate in questo lavoro di tesi.

1.1 Marcatori molecolari microsatelliti: caratteristiche ed utilizzo

I microsatelliti o SSR consistono in ripetizioni in tandem (da 10 fino a 30 unità ripetute) di corti motivi (da 1 a 6 bp), sparse nei genomi della maggior parte delle specie eucariotiche, con una frequenza di circa 10^4 - 10^5 per genoma (Tautz & Schlotterer 1994; Jarne & Lagoda 1996). Si possono distinguere come ripetizioni perfette, ovvero senza interruzioni tra le unità ripetute, ripetizioni imperfette perchè interrotte da nucleotidi non ripetuti e ripetizioni composite dove due o più motivi SSRs sono trovati l'uno adiacente all'altro.

I microsatelliti sono altamente variabili, è stato stimato un tasso di mutazione che va da 10^{-6} a 10^{-2} per locus per generazione, significativamente più elevato del tasso di mutazione spontanea che è nell'ordine di 10^{-8} - 10^{-9} per locus per generazione (Ellegren, 2000). Il tasso di mutazione ai loci microsatelliti varia tra i loci e tra gli alleli dello stesso locus.

La variabilità nel numero di unità ripetute è principalmente dovuto allo scivolamento della DNA polimerasi (replication slippage) durante i meccanismi di replicazione, là dove questi errori non siano corretti dal sistema enzimatico di riparazione dei mismatch, e in misura minore da fenomeni di ricombinazione ineguale (Levinson & Gutman 1987; Stand et al., 1993; Ellegren, 2004). I loci SSRs con più ripetizioni tendono a mutare maggiormente poiché lo scivolamento della polimerasi incrementa in proporzione al numero di unità ripetute. L'espansione dei microsatelliti è più frequente quando la sequenza di DNA che contiene il motivo ha dimensioni ridotte, viceversa la contrazione è più probabile quando il numero di ripetizioni è ampio (Lai & Sun 2003).

Sembra che queste sequenze ripetute siano mantenute solamente grazie alla loro abilità a replicarsi ed espandersi nel genoma all'interno comunque dei limiti dettati dalle pressioni selettive negative che agiscono sulla perdita di fitness che questi potrebbero causare (Charlesworth et al., 1994).

La disponibilità di risorse genomiche e librerie EST (express sequence tag) nelle piante ha permesso di valutare la loro abbondanza e distribuzione (Morgante et al., 2002). In generale i microsatelliti si ritrovano con maggior frequenza nelle regioni non codificanti, anche se i motivi tri- ed esa-

nucleotidici, contrariamente a quanto riportato in lavori precedenti sui genomi animali, hanno una frequenza più alta nelle regioni trascritte, specialmente nelle porzioni 5'-UTR, mostrando quindi un'associazione significativa con la frazione a basso numero di copie nel genoma delle piante (Morgante et al., 2002). Questa tipologia di marcatori SSR associati alle regioni trascritte, ha il vantaggio che può essere trasferita tra i vari taxa poiché le regioni fiancheggianti sono molto conservate ed evitare le problematiche relative allo sviluppo *ex-novo* di marcatori (Kashi et al., 1997).

I microsatelliti sono utilizzati, nel campo agrario-forestale, per un'ampia gamma di applicazioni come gli studi di genetica di popolazione, la caratterizzazione di germoplasma e l'analisi varietale, le analisi di paternità, la genotipizzazione e la costruzione di mappe genetiche, la tassonomia sistematica, l'evoluzione molecolare, la descrizione degli eventi di ibridazione (Morgante & Olivieri 1993). La valutazione della variabilità genetica all'interno delle specie perenni di interesse agrario ha importanti conseguenze nel miglioramento varietale e nella conservazione delle risorse genetiche. Uno dei principali utilizzi nel campo agrario è relativo alla caratterizzazione di germoplasma e l'identificazione varietale. Sono molti i lavori che utilizzano i marcatori SSRs su specie importanti nelle regioni Mediterranee tra le quali l'olivo (Bracci et al., 2009), la vite (Botta et al., 1995), il melo (Goulao & Oliveira 2001), il fico (Achtak et al., 2009). Questa tesi prende in esame un'altra specie importante nelle regioni Mediterranee, la palma da dattero (*Phoenix dactylifera*), la caratterizzazione della sua diversità genetica per scopi di identificazione varietale, mediante un sistema di marcatori SSR.

I microsatelliti nucleari con un contenuto polimorfico elevato sono stati utilizzati frequentemente negli studi di filogeografia, anche se i loro elevati tassi di mutazione possono rappresentare uno svantaggio data l'elevata omoplasia insita in questi marcatori (Doyle et al., 1998) che incrementa con gli anni di divergenza e la distanza genetica delle unità tassonomiche prese in considerazione (Provan et al., 2001; Jakob et al., 2007).

Per ridurre i possibili effetti dovuti all'omoplasia sono stati introdotti i marcatori SSR citoplasmatici (plastidiali e mitocondriali) per questi tipi di studi e recentemente sono stati proposti gli SNPs come marcatore

preferenziale per le analisi multi-locus delle popolazioni (Brumfield et al. 2003; Zhang & Hewitt 2003).

Uno dei principali problemi che ha limitato la diffusione dei microsatelliti, in particolare in specie dove le risorse genomiche sono scarse, riguarda gli aspetti relativi al loro sviluppo che sono spesso costosi sia in termini di tempo che di risorse economiche. Uno dei principali approcci utilizzati prevede la costruzione di librerie genomiche arricchite, l'identificazione, attraverso l'ibridazione di specifiche sonde, dei cloni che contengono la sequenza SSR, il sequenziamento di queste porzioni e il disegno di primer specifici. Successivamente i microsatelliti vanno caratterizzati per la possibilità di ottenere un contenuto informativo polimorfico (Edwards et al., 1996). La recente introduzione delle tecnologie basate sul sequenziamento di nuova generazione permettono di superare le problematiche relative alla costruzione di librerie e consentono di utilizzare metodi bioinformatici di ricerca *in silico* di motivi SSR, con innumerevoli benefici sia in termini di tempo che di costi (Duran et al., 2009). La procedura di screening è stata sviluppata appositamente per questi marcatori ed è automatizzata e ad alto processamento (Mitchell et al., 1997).

I marcatori SSR possono ritrovarsi, oltre che nei genomi nucleari, anche in quelli citoplasmatici (cpSSR e mtSSR), i primi sono codominanti mentre quelli che derivano dai genomi degli organelli sono aploidi. I loci cpSSR consistono principalmente di ripetizioni a singolo nucleotide e visto che il DNA plastidiale è altamente conservato è possibile, una volta identificato il motivo, disegnare coppie di primers trasferibili in differenti specie. Questi primer universali sono stati sviluppati per le Angiosperme dicotiledoni e monocotiledoni (Weising & Gardner 1999; Chung & Staub 2003) come per le Gimnosperme (Vendramin et al., 1996).

L'insieme effettivo dei loci plastidiali è circa la metà di quello dei loci nucleari a causa del loro stato aploide e della trasmissione uniparentale. Questa caratteristica rende i marcatori plastidiali più sensibili ai processi stocastici quali eventi di deriva genica ed effetto del fondatore. I cpSSR sono stati sfruttati per testare ipotesi sui processi di dispersione, per identificare le rotte di colonizzazione/migrazione post glaciali, per comprendere i fenomeni di differenziazione intra-specifica e l'introgresione inter-specifica (Provan et al., 2001; Petit et al., 2003; Cronn et al., 2008).

2.2 *Marcatore molecolari SNP: caratteristiche ed utilizzo*

Gli SNPs costituiscono la risorsa più abbondante di polimorfismi a livello genomico, e descrivono il cambiamento di una singola base ad uno specifico locus genomico dovute a mutazioni puntiformi (transizioni, transversioni e inserzioni/delezioni o indel). Gli SNP teoricamente potrebbero essere bi- tri- o tetra-allelici, anche se gli SNP tri- e tetra-allelici sono molto rari e per questo sono considerati generalmente marcatori bi-allelici (Doveri et al., 2008). Potrebbe sembrare uno svantaggio se si confrontano con marcatori multi-allelici come i microsatelliti, invece questo è compensato dalla loro abbondanza relativa, infatti, una sequenza di interesse potrebbe essere altamente densa di marcatori SNP. E' stato stimato su alcune linee di mais che uno SNP appare circa ogni 40-75 bp (Gore et al., 2009) mentre in alcune piante forestali appartenenti alla famiglia delle Pinaceae la frequenza è di circa uno SNP ogni 100 bp (Savolainen & Pyhajarvi 2007). I livelli di polimorfismo nelle piante sono 10-20 volte più elevati rispetto alla maggior parte delle specie animali.

Considerato che la maggior parte di essi sono biallelici e si ritrovano sparsi nel genoma a frequenze elevate i marcatori SNP sono appropriati per le metodiche ad elevata processività (Ranade et al., 2001). Questi marcatori molecolari hanno tassi di mutazione nell'ordine di 10^{-8} - 10^{-9} per locus per generazione, evitando le problematiche inerenti all'omoplasia che si hanno con i marcatori SSR soprattutto quando si adoperano per fare inferenze filogenetiche (Brumfield et al. 2003).

Gli SNPs si ritrovano più frequentemente nelle regioni non codificanti piuttosto che in quelle codificanti ma la frequenza media varia tra le specie.

Il ri-sequenziamento di porzioni genomiche, amplificate via PCR, è il modo più diretto per identificare gli SNPs. I primer sono disegnati per generare frammenti di dimensioni di 400-700 bp, frequentemente derivati da geni candidati o EST. Esistono diversi database EST disponibili per identificare i polimorfismi SNPs in alberi di interesse forestale (Neale & Kremer 2011). Un'altra alternativa per le specie in cui le risorse genomiche sono limitate, come quelle appartenenti alle conifere, può essere quella di ricercare SNP in geni orologi partendo da informazioni disponibili in specie modello (Brown et al., 2001).

Per ottimizzare l'identificazione di SNPs l'amplificazione di una sequenza candidata può essere eseguita su un gruppo di popolazioni altamente

divergenti, massimizzando ad esempio il gradiente ecologico della specie in esame, ognuna delle popolazioni rappresentata da più individui (Krutovskii & Neale 2001). Questa strategia, come descritto più in dettaglio nel capitolo 2, è stata utilizzata per individuare gli SNP in una specie forestale non modello, il pino nero (*Pinus nigra*).

L'avvento delle tecnologie di sequenziamento di nuova generazione e l'identificazione diretta degli SNPs *in silico* offre oggi un'efficiente strategia alternativa al ri-sequenziamento degli ampliconi per lo sviluppo degli SNPs nelle piante (Nordborg & Weigel 2008).

La visualizzazione dei siti eterozigoti è facilitata dall'utilizzo di specifici software di allineamento di sequenza e di controllo di qualità del segnale (Phred e BioEdit (Ewing et al., 1998)).

Nel caso specifico delle conifere, il DNA sequenziato dal megagametofito (il tessuto aploide di origine materna) elimina i problemi dell'identificazione dell'eterozigote e della ricostruzione dell'aplotipo, problema quest'ultimo, che può essere superato utilizzando alcuni software statistici specifici nel caso in cui i polimorfismi SNP fossero analizzati in un tessuto diploide.

La genotipizzazione dei marcatori SNPs sta diventando uno strumento essenziale per studi di genomica di popolazione, associazione genetica, caratterizzazione di germoplasma, miglioramento varietale, creazione di mappe genetiche ad alta risoluzione in svariate specie modello e non modello (Grattapaglia et al., 2011).

Gli SNP a livello dei geni candidati sono utilizzati principalmente negli studi di associazione genetica per analizzare caratteri complessi perché nelle specie arboree consentono di evitare la costruzione di vaste popolazioni segreganti come richiesto nelle classiche procedure di mappaggio di QTL. Il rapido decadimento del linkage disequilibrium (LD) atteso nelle popolazioni naturali dopo poche generazioni di incrocio casuale, consente di interpretare l'eventuale associazione tra marcatore e carattere di interesse come dovuta all'esistenza di elevata concatenazione o addirittura alla presenza del marcatore SNP nella sequenza codificante del gene responsabile della variante funzionale (Neale & Savolainen 2004). Questa caratteristica è fondamentale per le applicazioni di miglioramento assistito dai marcatori (MAS) visto che l'associazione marcatore-carattere di interesse è poco probabile che si perda a causa dei successivi cicli di riproduzione e ricombinazione. Questi studi hanno permesso di identificare associazioni per

diversi caratteri in molte specie (per alcuni esempi su specie arboree forestali vedi Neale & Kremer 2011) anche se l'associazione marcatore-carattere spiega solo una piccola porzione della varianza genetica.

L'identificazione di marcatori SNP all'interno di geni candidati di interesse adattativo (fenologia, tolleranza/resistenza a stress abiotici e biotici) sta diventando un approccio emergente per costituire studi di associazione molecolare in un definito contesto ecologico ed evolutivo, nei quali i fenotipi e le componenti ambientali sono dissezionati per permettere un'accurata analisi di correlazione tra fenotipi e genotipi (Eckert et al., 2009). Un esempio indicativo di questi approcci è rappresentato dagli studi della variazione clinale nelle specie forestali (Savolainen & Pyhäjärvi 2007). La maggior parte delle specie forestali, infatti, mostra una variazione continua per i caratteri fenologici lungo gradienti geografici ed alcune indagini sono state pianificate con lo scopo di condurre test di associazione simultanei per trovare indicazioni di variazione clinale a livello genomico (Ingvarsson et al., 2006). La variazione genetica adattativa delle popolazioni forestali e i corrispondenti dati ecologici possono essere combinati con modelli climatici per valutare l'abilità della specie a rispondere o con la migrazione o con l'adattamento a cambiamenti ambientali (Rehfeldt et al., 2002). La comprensione del potenziale genetico adattativo delle specie arboree forestali è di cruciale importanza per valutare i loro rischi associati con i cambiamenti climatici.

CAPITOLO 1

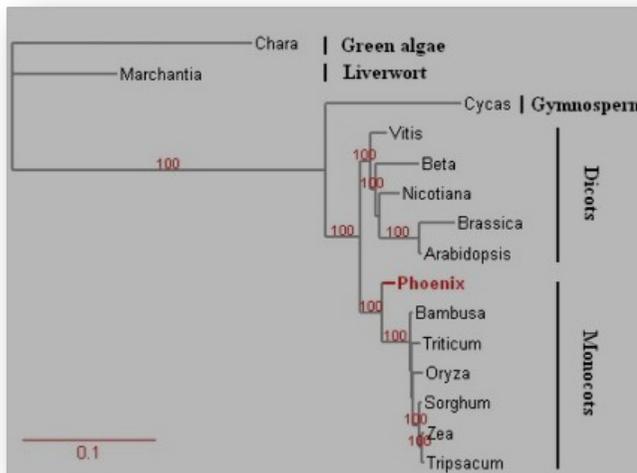
Analisi della diversità genetica della palma da dattero (*Phoenix dactylifera* L.) nell'Oasi di Al-Jufrah (Libia)

1. INTRODUZIONE

Le Palme o Arecaceae sono tra le poche famiglie di piante monocotiledoni che presentano specie longeve con portamento arboreo, essendo sprovviste di accrescimento secondario. Le Arecaceae costituiscono l'unica famiglia dell'ordine delle Arecales e ad oggi si conoscono 183 generi con circa 2500 specie la maggior parte delle quali diffuse nelle zone tropicali e sub-tropicali (Dransfield et al., 2008a; Horn et al., 2009). Si ipotizza che le palme siano una delle prime famiglie moderne delle monocotiledoni e questa tesi viene corroborata sia da studi di filogenetica molecolare che da ritrovamenti paleontobotanici. Infatti, studi in cui si sono confrontate le sequenze di geni mitocondriali di *P. dactylifera* con gli omologhi di altre 15 specie hanno evidenziato che questo genere *clusterizzava* alla radice della branca relativa alle monocotiledoni (Figura 1). Inoltre, sono numerosi i ritrovamenti fossili di palme datati a partire dal periodo del tardo Cretaceo (circa 80 milioni di anni fa) (Moore et al., 1982; Harley 2006; Fang et al., 2012). Vista la ricchezza di specie presenti, si ipotizza che questa famiglia abbia subito un breve periodo di radiazione adattativa, che ha portato, a partire da un antenato comune, a una rapida diversificazione di nuove specie, ognuna delle quali adattata ad occupare una specifica nicchia ecologica (Harley & Baker, 2001).

Questa famiglia comprende una serie di specie di importanza agro-economico-sociale come la palma da dattero, da cocco e da olio, inoltre molte specie tra i vari generi hanno rilevanza vivaistica sono, infatti, utilizzate come piante ornamentali e decorative. Queste specie, da un punto di vista della biodiversità e in ottica di conservazione della diversità genetica, sono tutte da salvaguardare.

Figura 3.1: Analisi filogenetica dedotta dal confronto di svariati geni in 15 genomi mitocondriali di piante (Fonte: Fang et al., 2012).



1.1 Introduzione sul genere *Phoenix*: tassonomia e filogenetica.

Il genere *Phoenix* (Arecaceae, Coryphoideae) include 14 specie dioiche (Tabella 1.1), con caratteristiche foglie pinnate, distribuite nelle latitudini tropicali e sub-tropicali dalle isole Canarie nell'oceano Atlantico, attraverso l'Africa e l'Asia, raggiungendo le isole Filippine (Govaerts & Dransfield, 2005). Il principale centro di diversità del genere si estende dall'India all'Indocina, dove sono presenti la maggior parte delle specie (Zaid & de Wet, 2002a). La condizione dioica è considerata come un sistema di riproduzione sessuale raro nelle piante a fiore, con solo il 4-6% di tutte le specie di piante angiosperme che la presentano (Renner & Ricklefs, 1995), e il genere *Phoenix* all'interno della famiglia delle Palme è uno dei pochi esempi. Da un punto di vista citologico tutte le specie analizzate presentano un genoma nucleare diploide formato da 36 cromosomi. Evidenze sperimentali ottenute utilizzando tecniche di ibridazione in situ, indicano che la determinazione del sesso è di tipo genetico (sistema X-Y) ed il cariotipo maschile è quello eterogametico, sebbene ancora non esista una mappa genetica o fisica che confermi la presenza dei cromosomi sessuali (Beal 1937; Siljak-Yakovlev et al., 1996). Questo genere è considerato come monofiletico, infatti, le barriere riproduttive sembrano essere molto limitate e sono molti gli esempi di ibridi interspecifici vitali (Nixon 1936; Moore 1963;

Henderson et al., 2006); questa tesi è inoltre corroborata sia da uno studio sulle sequenze plastidiali (Dransfield et al., 2008b) che dall'elevata trasferibilità, all'interno del genere, di un set di marcatori microsatelliti (SSRs) sviluppati appositamente per la palma da dattero (Bilotte et al. 2004).

Tabella 1.1: Specie accettate appartenenti al genere *Phoenix* e loro distribuzione geografica.

Specie	Distribuzione
<i>Phoenix acaulis</i> Roxb.	Bangladesh e India
<i>P. andamanensis</i> Barrow	Isole Andamane (India)
<i>P. atlantica</i> Chev.	Isole di Capo Verde
<i>P. caespitosa</i> Chiov.	Penisola Arabica
<i>P. canariensis</i> Chabaud	Isole Canarie (Spagna)
<i>P. dactylifera</i> L.	Nord Africa, Penisola Arabica, importata anche in California
<i>P. loureiroi</i> Kunth	Sub-continente Indiano, Cina meridionale, Filippine
<i>P. paludosa</i> Roxb.	India e Indonesia
<i>P. pusilla</i> Gaertn	India e Sri Lanka
<i>P. reclinata</i> Jacq.	Africa tropicale e Madagascar
<i>P. roebeleni</i> O'Brien	Indo-Cina
<i>P. rupicola</i> Anders.	India
<i>P. sylvestris</i> Roxb.	Subcontinente Indiano
<i>P. theophrasti</i> Greuter	Grecia del sud, Creta, Turchia

A livello morfologico la diversità è molto ristretta tra le differenti specie di *Phoenix*, con pochi descrittori capaci di discriminare in maniera non

ambigua tra specie strettamente correlate come *P. dactylifera* da *P. theophrasti* (Barrow, 1998) o *P. atlantica* da *P. dactylifera* (Henderson et al., 2006). Un esempio, che è stato motivo di numerose speculazioni da un punto di vista tassonomico, riguarda soprattutto il frutto, infatti, oltre alla *dactylifera* altre specie presentano datteri edibili (*P. acaulis*, *P. atlantica*, *P. loureiroi*, *P. reclinata*, *P. theophrasti*), e questo ha causato non pochi problemi di attribuzione della specie. A livello molecolare il genere *Phoenix* è difficile da collocare da un punto di vista filogenetico. Molte sono le ricerche che prendono in considerazione quelle regioni genomiche universali comunemente usate in studi filogenetici, sia nucleari (regioni intergeniche del 5S e 18S DNA) che plastidiali, ma nessuno di questi ha ancora risolto le relazioni che intercorrono tra le specie (Asmussen & Chase 2001; Hahn 2002).

Questi bassi livelli di variazione morfologica e genetica, hanno portato a sviluppare nuovi marcatori molecolari più polimorfici che chiarissero il pattern di variazione genetica all'interno del genere *Phoenix*. Un impulso fondamentale al raggiungimento di questo obiettivo è derivato dal lavoro di Bilotte e collaboratori (2004). Questi ricercatori hanno mostrato che la maggior parte dei marcatori microsatelliti sviluppati ex-novo per la palma da dattero (*Phoenix dactylifera*) sono facilmente trasferibili a tutte le altre specie di *Phoenix*, indicando stretta affinità interspecifica e permettendo un'indagine più accurata in questo genere. Henderson (Henderson et al., 2006) ha usato questo set di microsatelliti con lo scopo di distinguere *P. atlantica* da *P. dactylifera*, *P. sylvestris* e *P. canariensis*; in questo studio tutte e quattro le specie sono state chiaramente separate con l'individuazione di molti alleli specie-specifici. Inoltre, per la prima volta è stato stabilito con studi genetici che *P. atlantica* non è frutto dell'ibridazione tra *dactylifera* e *canariensis*, come si ipotizzava, ma bensì una “vera” specie. Più recentemente Pintaud e collaboratori (2010) utilizzando questi marcatori microsatelliti, accoppiati ad un locus minisatellite plastidiale altamente polimorfico sia all'interno che tra le specie (psbZ-trnfM, Henderson et al., 2006), hanno ottenuto una buona delimitazione tra le undici specie prese in esame permettendo un'analisi di clustering accurata. La diversità genetica valutata con questo set di marcatori SSRs è risultata elevata in quelle specie con ampia distribuzione (*P. dactylifera*, *P. loureiroi*, *P. reclinata*), mostrando svariati alleli polimorfici ad ogni locus considerato, e come atteso

bassa in quelle specie con distribuzione ristretta (*P. acualis*, *P. canariensis*, *P. rupicola*, *P. pusilla*, *P. rupicola*, *P. roebelenii*, *P. theophrasti*) con alleli privati fissati all'interno delle specie, un tipico pattern consistente con l'evoluzione di piccole popolazioni isolate. Discorso totalmente diverso va fatto per *P. atlantica*, un endemismo dell'isola di Capo Verde, che presenta una diversità genetica simile a quella di *P. dactylifera*, con diversi alleli condivisi. Il risultato è sorprendente se si pensa all'ampia distribuzione della *dactylifera* e spiegabile o con un campionamento non rappresentativo o con una depauperazione della diversità genetica della palma da dattero causata dalla selezione antropica e da specifiche epidemie, come la patologia del Bayoud responsabile della distruzione di migliaia di palme (Djerbi 1982). Entrambi i fattori non hanno influenzato *P. atlantica*, la cui elevata diversità può essere mantenuta attraverso una suddivisione della popolazione come risultato di un ristretto flusso genico attraverso l'isola di Capo Verde.

L'elevata omoplasia riscontrata in molti caratteri morfologici, il ridotto tasso di evoluzione dei marcatori molecolari standard unito alle scarse informazioni genetiche presenti in letteratura (Wilson et al., 1990; Gaut et al., 1992; Eyre-Walker & Gaut, 1997) limitano la chiara comprensione delle relazioni tra le specie. Con il sequenziamento ex-novo del genoma nucleare (Malek 2010) e plastidiale (Yang et al., 2010) della *P. dactylifera*, effettuato attraverso una piattaforma di nuova generazione (pirosequenziamento, Roche) che trae vantaggi dall'abbattimento dei costi e dei tempi, si sono gettate le basi per futuri progetti di sequenziamento di genomi del genere *Phoenix*. Il confronto tra queste ampie porzioni genomiche permetterà di giungere a risultati importanti e portare ulteriori conoscenze sulla genetica, la filogenetica e l'evoluzione di questo complesso e importante genere.

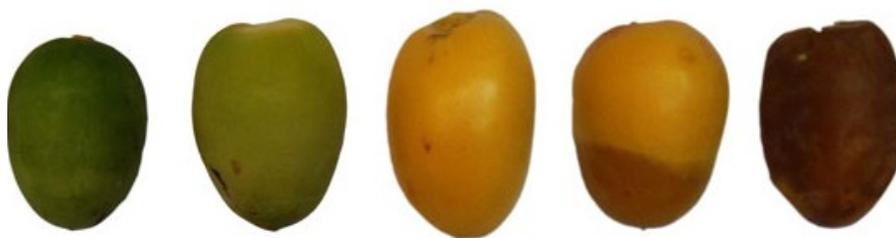
1.2 *Phoenix dactylifera*.

La palma da dattero (*Phoenix dactylifera* L.) con la sua storia di domesticazione e utilizzazione datata, sulla base di reperti di datteri fossili, tra i 7 e gli 8 mila anni fa (Zohary & Spiegel-Roy 1975) è considerata una specie cardine, soprattutto in quei sistemi agricoli desertici e predesertici del nord Africa e del Medio Oriente, dove condizioni particolari consentono la formazione dell'ecosistema oasi che permette la sopravvivenza delle popolazioni locali in questi ambienti estremi. Oltre all'inestimabile valore nutritivo del frutto, uno dei principali alimenti nella dieta di questi Paesi, la palma da dattero è una pianta multifunzionale che si utilizza anche per fibra,

combustibile e costruzioni (Chao & Krueger 2007). Questa specie dioica e con lungo ciclo vitale, in natura può riprodursi sia sessualmente (via seme) che asessualmente. La principale modalità di propagazione, nelle piantagioni commerciali, avviene per via vegetativa, vengono prelevati i polloni alla base della pianta madre, mantenendo l'integrità genetica della cultivar. Questo modo di propagazione agamica in vivo è anche l'unica alternativa, visto che, essendo una specie monocotiledone e quindi priva di un cambio vascolare, non può essere riprodotta per talee o innesti, a differenza di molte specie dicotiledoni. I polloni sono prodotti in numero limitato durante il ciclo vitale della palma, sia in relazione al tipo che all'età della cultivar, si va da pochi polloni fino a oltre 20 per pianta (Zaid and de Wet 2002b). Tale metodica ha i vantaggi rispetto alla riproduzione via seme di propagare una varietà già selezionata (alte rese, resistenza a patogeni e/o stress abiotici) e di eliminare la probabilità che metà dei semenzali sia di sesso maschile, cosa che sarebbe accertabile fenotipicamente solo dopo 5-8 anni quando compaiono le prime infiorescenze (Zaid & de Wet 2002a). Sebbene, con il sequenziamento del genoma di alcune cultivar di palma (Al-Dous et al., 2011) comincino a comparire in letteratura i primi lavori di messa a punto di protocolli per la distinzione del sesso utilizzando metodiche molecolari (Al-Mamhoud et al., 2012). La varietà maschile ancora oggi viene selezionata sulla base della quantità di polline prodotto. Un'ottima palma maschile dovrebbe essere sufficiente per impollinare più di 50 piante femminili, infatti, in una piantagione commerciale sono presenti individui maschili in rapporto di 5/50 individui femminili. E' stato dimostrato che il polline di varietà diverse può influenzare la dimensione del seme e del frutto così come il suo periodo di maturazione (Ream 1976). Diviene così importante riconoscere con esattezza anche le varietà maschili poiché i migliori risultati in termini qualitativi e quantitativi si ottengono usando il polline di una pianta della stessa varietà di quella da impollinare. In natura l'impollinazione è di tipo anemofilo, nei palmeti commerciali invece per massimizzare l'impollinazione e rese le cultivar vengono impollinate artificialmente. La palma da dattero raggiunge la sua piena maturità intorno ai 12 anni, e in media rimane produttiva per oltre 50 anni, nonostante siano stati descritti casi eccezionali dove la pianta protraeva la produzione oltre i 150 anni (Chao & Krueger 2007). Lo sviluppo del dattero è classificato in cinque stadi, in accordo con la nomenclatura araba: *Hababouk*, *Kimri*, *Khalal*, *Rutab*, *Tamar* (Figura 1.2). I frutti possono

essere raccolti e commercializzati negli ultimi due stadi di maturazione in funzione del tipo di cultivar e delle condizioni climatiche. La maggior parte delle varietà sono adatte ad essere consumate sia fresche (*Rutab*) che secche (*Tamar*), e sulla base della tessitura del frutto si possono raggruppare le cultivar in “soft”, “semidry” e “dry” (Zaid and de Wet 2002a). Il periodo della raccolta inizia ad agosto per le varietà precoci e si protrae fino a novembre per quelle che possono persistere sulla pianta, anche se i mesi prevalenti della raccolta sono settembre ed ottobre.

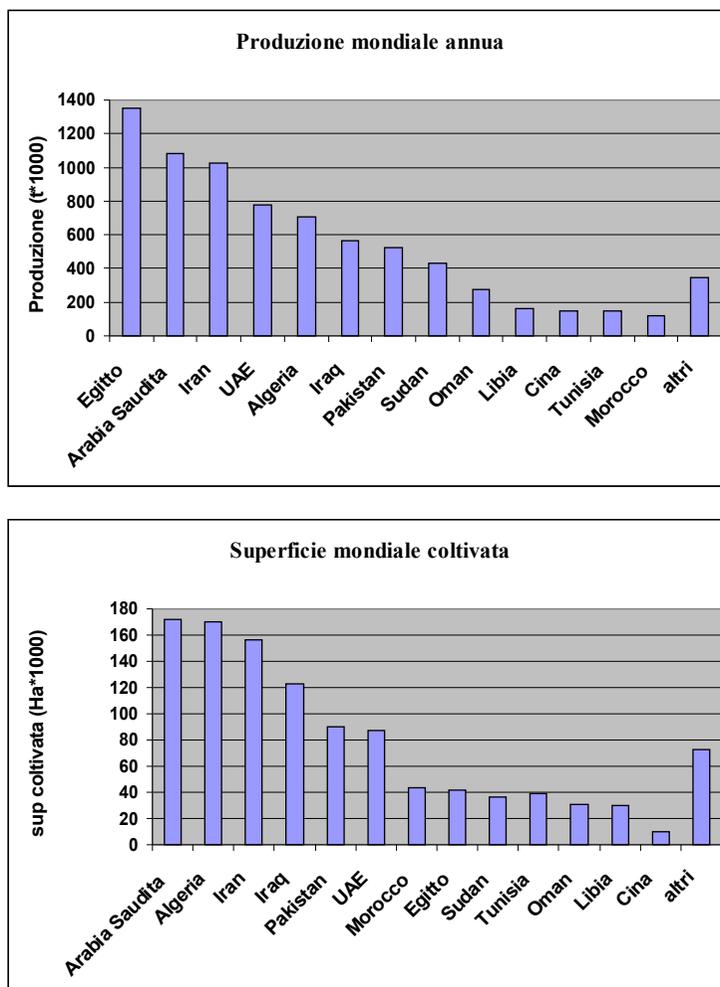
Figura 1.4: Rappresentazione delle 5 principali fasi di sviluppo del dattero (Fonte: Yin et al., 2012).



La fenicoltura in questi ultimi anni ha raggiunto dei valori importanti di produzione e superficie coltivata (Figura 1.3), basta fare riferimento agli ultimi risultati pubblicati dalla FAO per rendersi conto che occupa, soprattutto nei paesi Medio Orientali e Nord Africani, un'importante fetta del settore agro-economico. Nel 2010, a livello mondiale, la superficie coltivata ha raggiunto 1.1 milioni di ettari con una produzione di 7.6 milioni di tonnellate di datteri (FAOSTAT 2010). La coltivazione avviene principalmente nelle regioni aride del Nord Africa e del Medio Oriente con percentuali sul totale, rispettivamente, del 43 e 40% della produzione e 33 e 64% della superficie coltivata. E' stato stimato che oltre i 2/3 delle palme presenti in tutto il mondo siano localizzate nei paesi Arabi (FAO 1982). La maggior parte della produzione viene consumata all'interno dei paesi produttori come si evince dai dati pubblicati dalla FAO. Infatti, nel 2010, solo 850.000 tonnellate sono state esportate, di cui più del 90% dai paesi Medio Orientali e solo il 9% dai paesi del Nord Africa. L'Europa è uno dei principali importatori di dattero sia in termini di valore, con oltre 290 milioni

di dollari (il mercato mondiale si aggira intorno a 740 milioni \$), che di quantità, con oltre 120.000 tonnellate di datteri, la maggior parte dei quali provenienti da paesi produttori che si affacciano sul Mediterraneo (FAOSTAT 2010).

Figura 1.5: Produzione e superficie mondiale raggiunta dalla palma da dattero nel 2010 (Fonte: FAOSTAT 2010).



Il mercato importa quasi esclusivamente datteri della varietà *Deglet Nour*. L'elevata produttività, il rapido tasso di crescita unito alle caratteristiche organolettiche del frutto elevate, fanno di questa cultivar il modello principale di palma da dattero, soprattutto nei palmeti moderni del Maghreb.

Infatti, a partire dalla fine degli anni '70 in Marocco, Algeria e Tunisia, sotto la spinta dei governi centrali, è stato attuato un piano di rinnovamento dei vecchi palmeti con l'introduzione massiccia di questa varietà a discapito delle cultivar considerate meno pregiate, con percentuali che superano il 50% in questi paesi e oltre il 90% in alcune oasi Tunisine (Rhouma 1999). La coltivazione monovarietale, oltre alle problematiche note di erosione della diversità genetica della specie con una conseguente ridotta capacità di adattamento a fattori biotici e abiotici, è anche alla base del diffondersi di fitopatologie specifiche, più o meno caratterizzate, alcune delle quali sfociate in vere e proprie epidemie (Takrouni et al., 1988; Triky et al., 2006). La più nota è la patologia del Bayud, una fusariosi vascolare causata dal *Fusarium oxysporum* f. sp. *albedinis*, che nell'ultimo secolo ha distrutto oltre 10 milioni di palme in Marocco (Haddouch 1996) e più di metà nelle oasi dell'Algeria (Brac de la Perriere & Benkhalifa 1991). La progressione della malattia continua verso est minacciando le piantagioni tunisine, libiche ed egiziane. Inoltre è stato dimostrato che molte cultivar d'élite sono sensibili al Bayud, sebbene siano state identificate diverse varietà, povere da un punto di vista della qualità del dattero, tolleranti o resistenti alla fusariosi (Saaidi et al., 1981; Saaidi 1992). Oltre ai problemi fitosanitari e alla spiccata uniformità genetica delle piantagioni commerciali, altri fattori minacciano il pool genico di questa specie nei paesi del Maghreb e del Medio Oriente, quali la siccità che causa un'accelerazione dei processi di desertificazione e di salificazione delle acque di falda, e i lenti processi di rinnovamento delle piantagioni. Tutto ciò indica che svariate cultivar tradizionali potrebbero essersi perse a discapito della diversità genetica, evidenziando la necessità di sforzi urgenti per la caratterizzazione e conservazione del germoplasma esistente con il fine di preservare la biodiversità della specie.

Il germoplasma della palma da dattero è difficile da conservare e stoccare sotto forma di polloni o collezioni di campo e la crescente richiesta di varietà importanti per la messa a dimora di nuovi impianti produttivi, unito alla limitata capacità di produrre polloni della palma da dattero, ha contribuito in maniera essenziale allo sviluppo delle tecniche di coltura in vitro in questa specie (Tisserat 1979; El Hadrami & Baaziz 1995; Eke et al., 2005; Bekheet 2011). La micropropagazione, sebbene ampiamente utilizzata, non è ancora una tecnica efficiente in questa specie, infatti, in base al tipo di protocollo e al genotipo utilizzato, sono stati riportati casi di variazione somaclonale in

cui l'identità genetica della varietà si era persa. Anche se oggi, con l'ausilio dei marcatori molecolari, è possibile risalire al materiale di partenza e controllare precocemente il genotipo della cultivar propagata (Rival et al., 1998; Saker et al., 2000). La variazione somaclonale è usata per spiegare la comparsa di varianti genetiche derivanti da procedure di micropropagazione. Questa variabilità appare spesso nelle colture *in vitro* di tessuti come manifestazione dell'influenza epigenetica o cambiamenti nel genoma delle cellule indotte al differenziamento. E' tuttora utilizzata da numerosi gruppi di ricerca specializzati nel miglioramento varietale delle piante ornamentali, per trovare piante con mutazioni stabili per caratteri agronomici interessanti (Isabel et al., 1993; Soniya et al., 2001). Per quanto riguarda la palma da dattero, soprattutto a causa dei lunghi tempi di generazione e dell'assenza di mappe genetiche e fisiche che permettono approcci di selezione basati sui marcatori molecolari, non vi sono molte risorse genetiche a disposizione del miglioratore varietale. In passato alcuni ricercatori hanno tentato di utilizzare la variazione somaclonale per tentare di migliorare le varietà o crearne di nuove, ma con scarsi risultati (Larkin & Scowcroft 1981). L'unica risorsa degna di nota è stata ottenuta in California attraverso un programma di "back cross" finanziato dall'USDA e iniziato negli anni '40, che ha richiesto più di 30 anni di lavoro. Lo scopo principale è stato quello di produrre cultivar maschili e femminili superiori attraverso una strategia di reincrocio ripetuto di cultivar maschili su cultivar femminili selezionate, oggi sono al quinto-sesto ciclo di "backcross" per alcune cultivar (Barrett 1973). Un altro obiettivo di questo programma era quello di generare una collezione di semi da utilizzare nella propagazione ma i problemi legati all'identificazione precoce del sesso uniti alla ricombinazione genetica ne ha limitato l'utilizzo (Nixon & Furr 1965). Oggi, lo sviluppo di marcatori specifici per la determinazione del sesso e per la caratterizzazione varietale, consente di stabilire la similarità di una collezione di semi con la pianta madre (Al-Dous et al., 2011). La preservazione in centri di conservazione del germoplasma, di embrioni somatici, meristemi e cellule delle principali varietà, sta diventando un importante strumento per stoccaggi o il mantenimento del germoplasma a lungo termine, con importanti implicazioni nella produzione massiva di varietà e nel miglioramento delle varietà stesse. Per raggiungere l'obiettivo di conservazione del germoplasma della palma da dattero vanno fatti ancora molti sforzi per la genotipizzazione varietale che permetta

l'identificazione delle cultivar distribuite nelle principali zone di coltivazione.

1.3 Caratterizzazione varietale.

La palma ha un'ampia plasticità di adattamento a condizioni pedologiche e climatiche difficili, garantita dall'evoluzione e differenziazione di numerosissime varietà localmente adattate, che sono state sottoposte a pressioni selettive durante la millenaria storia di coltivazione in questi areali estremi. Sulla base dei descrittori morfologici sono state identificate più di 2000 varietà distribuite nei principali centri di produzione, che differiscono per forma e dimensione del frutto, colore, gusto e tempo di maturazione, morfologia della spata e della spina (Al-Farsi & Lee 2008). Quando la ricchezza delle varietà è valutata attraverso i descrittori pomologici, l'elevata plasticità della specie può contribuire a un'identificazione ambigua o erronea della varietà provocando distorsioni nella stima del germoplasma complessivo. Un'altra fonte di errore, come conseguenza degli antichi processi di scambio di materiale vegetale tra varie regioni o Paesi, deriva dalla nomenclatura. Infatti, la stessa cultivar potrebbe avere nomi diversi o vice versa cultivar diverse essere classificate come uguali nei vari Paesi di produzione del dattero (Torres & Tisserat 1980; Cao & Chao 2002). I caratteri morfologici, sebbene siano indicatori variabili o imprecisi, sono stati usati spesso in passato per descrivere e identificare le varietà di palme da dattero del Nord Africa del genotipo della cultivar poiché influenzati dall'ambiente e dagli stadi di sviluppo della pianta (Elhoumaizi et al., 2002; Salem et al., 2008). I marcatori molecolari, soprattutto quelli basati sui polimorfismi a livello di DNA, sono degli strumenti idonei per la genotipizzazione varietale e in letteratura esistono diversi lavori di caratterizzazione del germoplasma fenicolo Nord Africano che utilizzano questo tipo di analisi. Gli strumenti molecolari utilizzati in questi lavori cambiano sulla base dell'importanza relativa di quel marcatore in quel determinato periodo, nonché dallo stato di avanzamento tecnologico del Paese in cui la ricerca è stata effettuata. I dati di variazione genetica trovati utilizzando marcatori come i Random Amplified Polymorphic DNA (RAPD) sono stati utilizzati per la genotipizzazione varietale di alcune accessioni del Marocco (Sedra et al., 1998), della Tunisia (Trifi et al., 2000) e dell'Egitto (Soliman et al., 2003; Adawy et al., 2006). Come riportato da alcuni autori,

questo tipo di marcatori non consentono una distinzione delle varietà, mostrando pattern monomorfici quando si analizzano cultivar affini. Altri marcatori come gli Amplified Fragment Length Polymorphism (AFLP) sono stati usati per valutare la diversità genotipica delle palme tunisine (Rhouma et al., 2007, 2011), egiziane (El-Assar et al., 2003, 2005) e californiane (Cao & Chao 2002). Questo tipo di marcatore pur mostrando un elevato polimorfismo nelle accessioni analizzate è una tecnica laboriosa sia in termini di sviluppo che di interpretazione dei risultati. Una maggiore comprensione dell'organizzazione genetica del germoplasma della palma da dattero è stata ottenuta utilizzando marcatori microsatelliti (SSRs). Grazie alla loro natura codominante e all'elevato numero di alleli per locus sono considerati ottimi candidati per la genotipizzazione degli individui e/o dei cloni nonchè per valutare gli errori di propagazione sia all'interno delle aziende che delle collezioni (Dangl et al., 2001). Fino al 2010, anno in cui è uscito la prima bozza della sequenza del genoma della palma, erano presenti solo due lavori di sviluppo ex-novo di marcatori SSR, ottenuti utilizzando librerie arricchite in motivi dinucleotidici (GT)_n e (GA)_n (Billotte et al., 2004; Akkak et al., 2009). Questi marcatori microsatelliti sono stati utilizzati per valutare la relazione e la diversità genetica delle varietà di palma in Mauritania (Bodian et al., 2012a), Marocco (Bodian et al., 2012b), Tunisia (Zehdi et al., 2004a, b; Hammadi et al., 2011), Algeria (Akkak et al., 2009), Sudan (Elshibli & Korpelainen 2008), Oman (Al-Ruquaishi et al., 2008), Qatar (Ahmed & Al-Qaradawi 2009) e California (Johnson et al., 2009). Uno studio del 2012 mette in luce la possibilità di utilizzare questo tipo di marcatori anche per la discriminazione, nelle fasi precoci, del sesso degli individui (Elmeer & Mattat 2012).

1.4 Sequenziamento e caratterizzazione del genoma di P. dactylifera.

Il sequenziamento del genoma ha permesso la sua esplorazione attraverso strumenti bioinformatici specifici per la ricerca di motivi SSR funzionali (Hamwiah et al., 2010). I risultati indicano che i motivi SSRs più abbondanti sono quelli con ripetizioni di-nucleotidiche (oltre 52000), seguono quelli formati da ripetizioni tri-nucleotidiche (oltre 28000 motivi) e penta-nucleotidiche (oltre 12800 motivi). La frequenza delle ripetizioni tetra-nucleotidiche sparse attraverso il genoma risulta molto più bassa rispetto agli altri tipi di ripetizioni (oltre 5000 motivi). Tra le ripetizioni di-nucleotidiche il tipo più comune è rappresentato da (GA)_n (48%) seguito da (AT)_n (37%). Questo lavoro ha permesso di disegnare oltre 1000 nuovi marcatori microsatelliti utilizzabili principalmente per la creazione di mappe genetiche più accurate nonché per l'analisi della diversità genetica, almeno per quel set di marcatori con un contenuto polimorfico adeguato (Elmeer et al., 2011). In un lavoro successivo sono state analizzate, mediante software specifici, le ripetizioni presenti nei geni “full-length” e sono stati individuati 2043 SSR con motivo ripetuto maggiore o uguale a 20bp. In accordo con i risultati ottenuti sull'intero genoma, le ripetizioni di-nucleotidiche (48%) sono le più abbondanti, seguite dalle ripetizioni tri- (42%) tetra- (6,5%) e penta-nucleotidiche (1,5%). I ricercatori, inoltre, hanno notato una forte distorsione tra regione CDS (codificante) e UTR (regione non tradotta), infatti, il 55% degli SSR tri-nucleotidici sono associati alle regioni codificanti e oltre il 90% dei microsatelliti di-, tetra- e penta-nucleotidici si ritrovano nella regione 5' e 3' UTR (Zhang et al., 2012) Questo lavoro potrebbe rappresentare un punto di partenza per lo sviluppo di marcatori SSR associati a geni che controllano importanti caratteri quantitativi utilizzabili sia nel campo del miglioramento genetico, attraverso tecniche di selezione assistita da marcatori (MAS), che in studi più generali legati a tematiche come adattamento e domesticazione. Per quanto riguarda la ricerca di marcatori SNP, prima di tutto sono stati confrontati gli alleli parentali della varietà di riferimento sequenziata (Al-Dous et al., 2011). Attraverso l'utilizzo di specifici software sono stati individuati oltre 1,7 milioni di SNPs in 381 Mb di sequenza ottenendo un tasso di eterozigosi dello 0,46% o uno SNP ogni 217 bp. La distribuzione è significativamente distorta e suggerisce la presenza di isole di elevato polimorfismo all'interno del genoma. Di questi SNPs circa 100.000 sono trovati all'interno di una sequenza predetta come

codificante e quasi 54.000 causano un cambiamento amminoacidico con un rapporto tra SNP non sinonimi e sinonimi di 1,17 molto simile al rapporto di 1,2 ritrovato in un'altra specie monocotiledone come il riso (McNally et al., 2009).

Per caratterizzare al meglio i polimorfismi nella palma da dattero sono state sequenziate, con differenti livelli di copertura del genoma, altre otto varietà, sia femminili che maschili, alcune non importanti dal punto di vista commerciale. Sono stati identificati oltre 3,5 milioni di SNPs polimorfici in almeno uno dei genomi sequenziati. Attraverso un algoritmo decisionale gli autori hanno identificato un “core” di 32 SNPs altamente informativo, con una minima perdita della potenza discriminativa rispetto a quando si utilizzano tutti i 3,5 milioni di SNP identificando tra questi cinque SNP il cui uso consente di distinguere tra le nove varietà sequenziale.

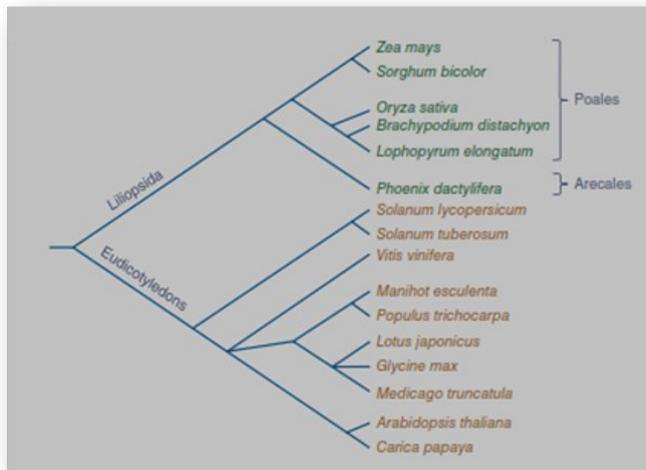
Questo set di marcatori SNPs fornisce un punto partenza, anche nella palma da dattero, per lo sviluppo di metodiche ad elevata processività per la genotipizzazione del germoplasma (Deulvot et al., 2010). Analisi recenti che si basano sulla citometria di flusso e sul confronto dello scaffold della bozza genomica contro una “batteria” fosmidica contenente l'intero genoma, suggeriscono una dimensione predetta del genoma nucleare della *dactylifera* di ~ 658 Mb. Il draft genomico sequenziato è stato calcolato essere di ~380 Mb, manca quindi approssimativamente il 40% del genoma totale, principalmente costituito da regioni ripetitive (Al-Dous et al., 2011). La figura 1.4 mostra un'albero filogenetico che comprende tutte le piante di importanza agronomica per le quali esistono, nelle banche dati, i loro genomi sequenziali. La palma da dattero è il primo membro delle Arecales e della famiglia delle Arecaceae per le quali è disponibile la sequenza del draft genomico. Esistono altre piante monocotiledoni con sequenze genomiche disponibili ma solo piante erbacee.

Pur essendo una delle piante con la storia di domesticazione più antica, agli inizi del 2010 erano presenti poco più di 100kbp di sequenze genomiche depositate in GeneBank e oltre al draft genomico in questi ultimi anni è stato sequenziato sia il genoma plastidiale, per una lunghezza di ~158Kbp (Yang et al., 2010) che quello mitocondriale, che è stato assemblato in una molecola circolare di 715 Kbp in lunghezza (Fang et al., 2012). Per quanto riguarda l'annotazione genica, sono state generate varie librerie a cDNA, sequenziate poi con piattaforme NG (next-generation), provenienti da vari

tessuti (foglie giovani e mature, radice, frutto, fiori e polloni) sia da piante femminili che maschili. Attraverso una specifica pipeline sono state predette le ORFs (open reading frame) per tutti i modelli genici per un totale che supera le 67500 predizioni, di cui il 73% con una ORF superiore ai 100 amminoacidi. Le ORF predette *in silico* sono state confrontate con i dati proteici presenti nel database Plant UniProt identificando 30850 modelli genici con porzione codificante intera e sono state validate *in silico* attraverso il mappaggio sul genoma assemblato di *P.dactylifera*. L'annotazione funzionale è stata eseguita allineando tutti i modelli genici contro vari database (NCBI e Plant UniProt) anche basandosi sui database di Gene Ontology (GO) e dell'Enciclopedia di Kyoto sui Geni e Genomi (KEGG). La distribuzione delle sequenze ortologhe è stata valutata in relazione a quattro genomi annotati, due monocotiledoni (*O.sativa* e *R. bicolor*) e due dicotiledoni (*A. thaliana* e *V. vinifera*). I modelli genici di *P. dactylifera* hanno condiviso il 70% (47.930 sequenze) e 69,4% (47.089) con gli ortologhi di riso e sorgo e il 68% (46.441) e il 69,3% (47.048) con le due dicotiledoni. E' stato trovato, inoltre, che 14880 modelli genici non allineano con i genomi di riferimento, e potrebbero rappresentare un "core" specifico della palma da dattero. Saranno necessarie ulteriori indagini a conferma del dato poiché le analisi basate sull'omologia sono sempre limitate dai parametri selezionati (Zhang et al., 2012).

Per quanto riguarda gli studi trascrittomici, il primo lavoro è stato eseguito confrontando gli mRNA del frutto della palma da olio e quella da dattero (Bourgis et al., 2011); successivamente, Yin e collaboratori (2012) hanno sequenziato il trascrittoma del frutto in sette stadi di maturazione, individuando una serie di geni associati alla fruttificazione, i loro livelli di espressione durante le varie fasi e le intricate attività metaboliche durante lo sviluppo e la maturazione del dattero.

Figura 1.6: Albero filogenetico delle più importanti specie agronomiche per le quali è disponibile una bozza genomica sequenziata. In verde sono rappresentate le specie monocotiledoni (Fonte: Al Dous et al., 2011).



1.5 Domesticazione della palma da dattero.

La domesticazione è un complesso processo evolutivo nel quale l'uomo induce cambiamenti morfologici e fisiologici alle specie animali e vegetali che porta a distinguere nettamente i taxa domesticati dai loro antenati selvatici (Hancock 2005). Sebbene la comprensione della domesticazione nelle piante annuali (pomodoro, girasole, riso, mais, grano) sia abbastanza avanzata grazie a recenti analisi archeologiche, genetiche e genomiche (Doebley et al., 2006; Glèmin & Bataillon 2009), ancora scarsa è la comprensione del modo in cui le piante perenni rispondono alla selezione artificiale (McKey et al., 2010). E' stato dimostrato, confrontando specie selvatiche con le loro affini domesticate, utilizzando marcatori molecolari neutri che i colli di bottiglia dovuti alla domesticazione non sono stati così intensi come nelle piante agrarie annuali; infatti, in quelle specie da frutto perenni antiche, come l'olivo, la vite e il melo, si è mantenuto in media il 95% della diversità genetica presente nei loro antenati selvatici (Lumaret et al., 2004; Aradhya et al., 2003; Coart et al., 2003). Questa larga porzione di variazione genetica totale è probabilmente il risultato di una combinazione di fattori, tra i più importanti vi sono: svariate popolazioni antenate distinte da un punto di vista geografico e genetico; fenomeni di ibridazione (inclusi gli ibridi generati dalle specie coltivate, il flusso genico mediato dall'uomo e il

flusso genico accidentale con le popolazioni natali sia antenate che non); simpatria congenerica. Comprendere l'origine geografica delle piante perenni da frutto potrebbe far luce sui primi stadi di domesticazione perché pochi cicli di riproduzione sessuale sono avvenuti tra specie perenni domesticate e i loro progenitori selvatici. E' stato dimostrato che tra le 18 specie perenni per le quali sono disponibili dati, 12 hanno un'origine multipla di domesticazione (Miller & Gross, 2011). Un esempio di una specie perenne mediterranea con origine multipla è l'olivo, una delle più antiche e meglio studiate specie domesticate. La pianta di olivo selvatica è nativa dell'area Mediterranea (Terral et al., 2004), ma individuare l'origine geografica precisa delle popolazioni coltivate è impresa difficile. Alcuni ricercatori hanno suggerito un'origine ovest e che gli olivi della zona est rappresentassero delle forme ferali (Bronzini de Caraffa et al., 2002). Dati di genetica molecolare hanno evidenziato almeno due origini dell'olivo coltivato, una zona ad est e l'altra zona ad ovest (Besnard & Bervillè 2000), e un altro studio ha identificato addirittura sette origini geografiche (Breton et al., 2006). Studiare gli aspetti della domesticazione e dell'origine geografica di una specie è un'impresa ardua soprattutto se le specie in questione sono state domesticate migliaia di anni fa e le informazioni genomiche sono scarse. L'origine e la storia di domesticazione della palma da dattero rimane ancora oscura. La maggiore complicazione risiede nel fatto che la diversità e la struttura del pool genico della palma da dattero è stato formato e fortemente alterato dalla selezione naturale e antropica, dalla propagazione clonale e dagli scambi e movimenti di germoplasma nello spazio e nel tempo ed è estremamente difficile dimostrare se questo taxon esista veramente in natura. Alcuni ricercatori hanno citato alcuni presunti popolamenti selvatici di *Phoenix dactylifera*, ma potrebbero benissimo essere popolazioni coltivate un tempo e oggi abbandonate (Zohary & Spiegel-Roy 1975; Zohary & Hopf, 2000). Sono state formulate diverse ipotesi per spiegare l'origine di *P. dactylifera*. Una prima ipotesi prevede che la palma da dattero, essendo una pianta fortemente coltivata, sia derivata per selezione antropica da una specie selvatica nota. *Phoenix sylvestris* è stata spesso vista come il possibile antenato selvatico della palma da dattero, come anche *P. atlantica* e *P. canariensis* (Munier, 1974). Un'altra ipotesi è che *P. dactylifera* potrebbe essere il risultato di un processo di ibridazione tra due o più specie selvatiche. Le presunte specie parenti che vengono menzionate in letteratura

includono *P. sylvestris*, *P. atlantica*, *P. canariensis* e anche *P. reclinata* (Popenoe, 1973). La terza e ultima ipotesi presume che *P. dactylifera* sia esistita o esista tutt'ora in natura come specie distinta nelle zone della penisola Arabica e nel Medio Oriente e sia stata domesticata indipendentemente dalle altre specie (Barrow, 1998). Quest'ultima ipotesi sembrerebbe essere confermata dai lavori molecolari di Henderson e Pintaud (Henderson et al. 2006 e Pintaud et al. 2010) che, basandosi dai risultati ottenuti da studi con marcatori microsatelliti, hanno identificato numerosi alleli privati suggerendo che *P. dactylifera* sia stata domesticata da popolazioni selvatiche della stessa specie, tuttavia ulteriori analisi sono necessarie per giungere ad una conclusione sicura.

La maggior parte degli studi molecolari sono in accordo con un'unica origine di domesticazione della palma da dattero avvenuta in Mesopotamia (Wrigley 1995). Tuttavia, sulla base di studi paleontobotanici e di ricchezza delle cultivar, altri ricercatori ipotizzano che la diffusione della palma da dattero e della sua coltivazione abbia seguito due distinte direzioni: una partendo dalla Mesopotamia all'Iran raggiungendo il Pakistan; l'altra partendo dall'Egitto attraverso la Libia raggiungendo quindi tutti i paesi del Maghreb (Zaid & de Wet, 2002b). Questa tesi sembrerebbe essere confermata da Pintaud (in prep) che valuta il polimorfismo di un minisatellite plastidiale nel genere *Phoenix*, trovando due sole varianti in *P. dactylifera* che sembrerebbero essere in accordo con le due origini. Anche questa tesi ha bisogno di ulteriori indagini e studi analizzando la palma da dattero in tutto il suo areale di distribuzione.

1.6 La fenicoltura libica.

La Libia italiana è stata una colonia del Regno d'Italia per più di 30 anni. In questo periodo furono condotte numerose le ricerche sulla fenicoltura libica ad opera degli studiosi del Ministero delle Colonie. Fin dai primi anni emerse subito la grande ricchezza varietale del germoplasma fenicolo del paese. In Libia, sulla base dei descrittori morfologici, sono state censite oltre 400 varietà e 95 di queste sono importanti da un punto di vista commerciale. C'è una grande differenza tra le varietà di dattero coltivate sulla costa e quelle coltivate nelle aree più interne, poiché il clima della costa per molte varietà non raggiunge le temperature necessarie per la completa maturazione del frutto (De Cillis 1923; Scarin 1937). Attualmente la coltivazione della palma da dattero in Libia presenta una situazione molto particolare. La

chiusura del Paese per le note vicende politiche dagli anni '70 in poi, se da una parte ha impedito lo sviluppo economico limitando notevolmente la diffusione delle varietà commerciali, dall'altra ha permesso di mantenere la coltivazione di moltissime varietà locali.

Le aree di coltivazione della palma da dattero in Libia possono essere suddivise, sulla base delle diverse latitudini, in tre zone principali (Figura 1.5):

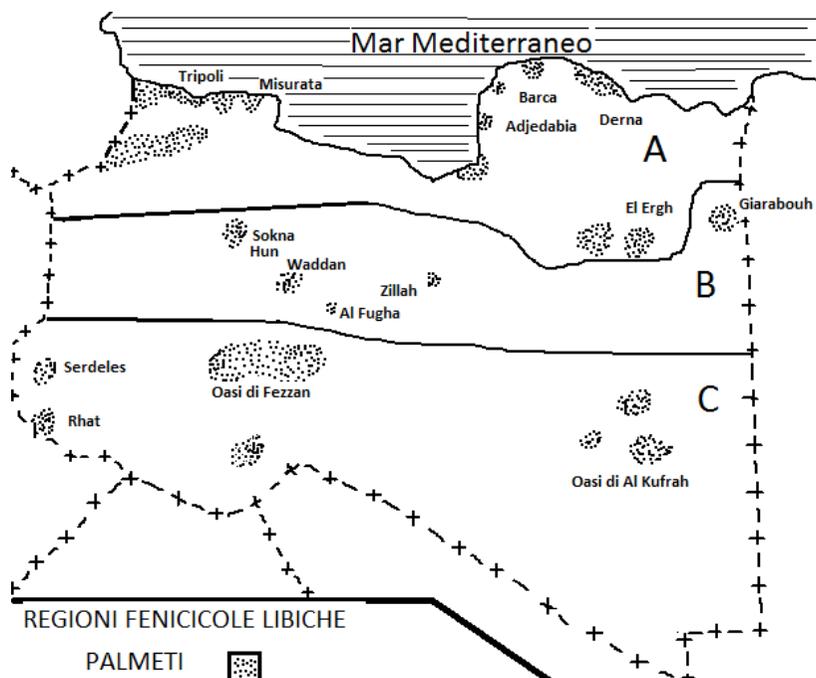
a) L'area costiera, situata a nord del 32° parallelo. In questa zona i migliori palmeti si trovano nelle aree di Tripoli, di Zuara, di Misurata e di Hisha dove l'acqua di irrigazione non scarseggia e le tecniche agronomiche sono avanzate. Tuttavia, nelle zone costiere si producono datteri di scarsa qualità, poco zuccherini e molto acquosi, a causa di temperature e umidità non ottimali.

b) L'area centrale, situata tra il 30° e il 28° parallelo. In questa zona le condizioni pedoclimatiche più favorevoli alla fenicoltura si trovano nelle aree pre-desertiche che corrono lungo la linea del 29° parallelo. Le località di Sokna, Hun, Waddan, Zella, Al Fugha, che insieme formano l'oasi di Al-Jufrah, rappresentano un'area fenicola di grande rilievo che conserva una notevole ricchezza varietale anche grazie al fatto che in passato erano tappe obbligate lungo le principali rotte carovaniere che collegavano la costa atlantica all'Egitto ed il Medio Oriente e quindi luoghi di scambi e mercati.

c) L'area meridionale situata al di sotto del 29° parallelo. La presenza della palma in questa zona è ristretta solo a una serie di oasi shariiane, tra le più note vi sono quella del Fezzan e di Kufra. I datteri presenti in quest'area appartengono alla qualità secca (dry).

Già negli anni del colonialismo i vari studiosi avevano individuato l'area centrale della Libia come sito ottimale per la crescita della palma da dattero e avevano ravvisato diverse varietà interessanti; infatti, un apposito decreto che ne disciplinava il commercio precisava agli agricoltori dell'oasi di Al-Jufrah che i datteri destinati all'esportazione erano solo quelli delle varietà *Abel*, *Bestian*, *Halima*, *Hamria*, *Kathari*, *Saiedi*, *Tagiat* e di questi solamente *Kathari* ed *Saiedi* erano considerati di prima qualità.

Figura 1.5: Principali zone di coltivazione della palma da dattero in Libia (Fonte: Munier 1978).



Nell'oasi di Al-Jufrah la maggior parte delle aziende produttrici di dattero è di piccole dimensioni e a conduzione familiare, sono presenti non più di 150-200 individui di varietà non sempre nota agli agricoltori locali, con la presenza di 3-4 varietà maschili ignote che servono come impollinatori principali. La maggior parte delle cultivar sono propagate attraverso i polloni sebbene alcune varietà di importanza secondaria siano state propagate per seme. Grazie all'isolamento geografico della conca di Al Jufrah, gli attacchi parassitari e fungini sono molto limitati. Contrariamente a quanto praticato negli altri Paesi del Maghreb e del Medio Oriente, nei nuovi palmeti artificiali realizzati negli ultimi anni in Libia si fa largo uso di varietà locali mentre ancora limitato è l'impiego di cultivar ad alta resa produttiva come la cultivar *Deglet Nour*.

Nell'ultimo decennio il Governo ha dato avvio ad un ampio programma di rilancio della fenicoltura e dell'olivicoltura nel paese. L'oasi di Al Jufra è

stata scelta per un intensivo programma di rinnovamento ed ampliamento dei palmeti supportata da una rete di distribuzione dell'acqua geologica derivata dai giacimenti petroliferi. A seguito degli ultimi rilievi effettuati nell'anno 2010 dall'Unità GIS dello IAO all'interno del progetto “*Miglioramento e valorizzazione della palma da dattero nelle oasi di Al Jufrah in Libia*” si stima che il numero di palme presenti nell'oasi di Al-Jufrah supera i 3 milioni.

2. SCOPO DELLA TESI

I principali obiettivi del lavoro di tesi sono:

- a) la caratterizzazione genetica con l'ausilio di marcatori microsatellite (SSRs) nucleari, di 18 cultivar campionate nell'oasi di Al Jufrah- Libia, con il fine di ottenere un profilo allelico caratteristico che consenta l'identificazione di ciascuna cultivar.

- b) valutare sulla base dei profili identificativi delle cultivar la variabilità esistente sia all'interno delle aziende di una stessa località che tra le aziende delle differenti località di campionamento e stimare l'accuratezza della propagazione clonale nonché la capacità degli agricoltori locali nel riconoscere le varietà.

- c) analizzare le cultivar libiche per un minisatellite plastidiale, associato sulla base dell'aplotipo all'origine, occidentale vs orientale, della palma da dattero allo scopo di valutare la distribuzione dei due aplotipi nel materiale fenicolo dell'oasi di Al-Jufrah.

- d) attuare un "fingerprinting" delle palme maschili propagate per seme allo scopo di potere effettuare una attribuzione ad una cultivar definita. Questa operazione è di fondamentale importanza in una prospettiva futura di miglioramento della qualità del frutto attraverso piani di incrocio controllati.

3. MATERIALI e METODI

3.1 Descrizione del materiale campionato.

Il campionamento del materiale vegetale è stato eseguito nell'oasi di Al Jufrah, un' ampia regione che si estende per circa 117.000 km² situata nella zona predesertica lungo il 29° parallelo Nord. La regione settentrionale di Al Jufrah è caratterizzata dalla presenza di una falda freatica e una artesia che hanno determinato la formazione di cinque insediamenti più o meno estesi: i tre principali sono Sokna, Hun e Waddan, e Al Fugha e Zella, ad una distanza di circa 200 km a sud-est dal nucleo principale (Figura 3.1).

Figura 3.1: Localizzazione dell'oasi di Al-Jufrah



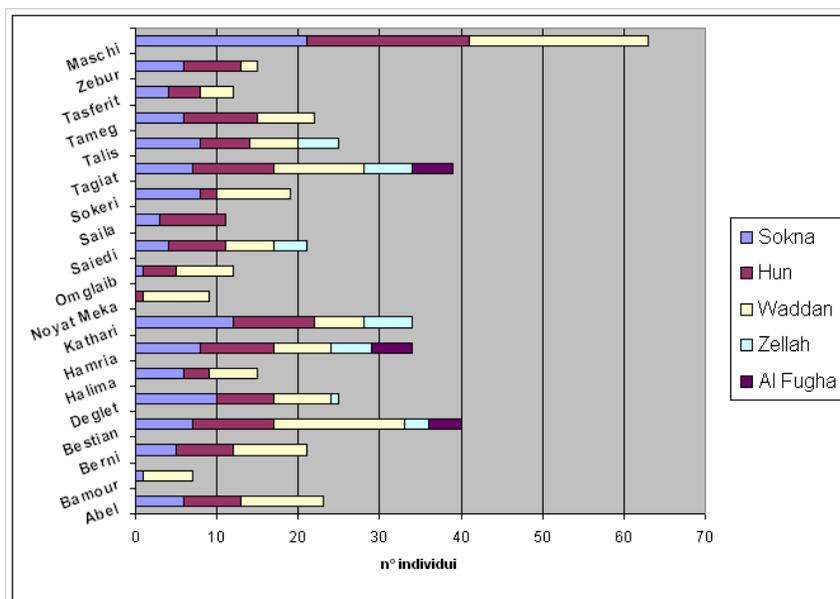
L'identificazione delle principali varietà di palma da dattero coltivate nell'area di Al-Jufrah è avvenuta con l'ausilio degli agronomi locali con i quali sono stati concordati i seguenti criteri di selezione: varietà autoctone tradizionalmente rilevanti, diffusione nella regione, importanza economica. Si sono individuate 18 varietà, precisamente: *Abel*, *Bamour*, *Berni*, *Bestian*, *Deglet*, *Halima*, *Hamria*, *Kathari*, *Noyat Meka*, *Omglai*, *Saiedi*, *Saila*, *Sokeri*, *Tagiat*, *Talis*, *Tameg*, *Tasferit*, *Zebur*. Sono state inserite nello studio anche due varietà identificate come non di origine locale, ma importanti in questa regione da un punto di vista economico: la *Deglet* (di origine algerina) e la *Saiedi* (di origine egiziana). Sono stati campionati, nelle cinque località, 484 piante femminili appartenenti alle 18 varietà, distribuiti in piantagioni, sia di recente che antica costituzione e 63 piante maschili di varietà ignota,

molti dei quali utilizzati come principali impollinatori nelle 38 aziende considerate. Le analisi sono state eseguite su un campione rappresentativo di 377 piante femminili e tutte e 63 le piante maschili (Figura 3.2); per una percentuale di copertura dell'83% a Sokna, del 78% a Hon, del 77% a Waddan, del 97% a Zillah e del 100% ad Al-Faqqah. In ciascuna azienda, sono state raccolte dall'apice vegetativo della pianta adulta le foglie giovani, scegliendo in ogni azienda almeno due-tre piante possibilmente di età diversa per ogni varietà. E' stato attribuito ad ogni pianta un numero di collezione composto dal codice dell'azienda, il codice della varietà e il numero dell'individuo. La figura 3.2 e la tabella 3.1 presentano, rispettivamente, la distribuzione del campionamento nelle cinque località delle varietà selezionate che rispecchia anche il tasso di diffusione di una data varietà nelle località e la distribuzione degli individui analizzati in funzione della varietà e località oggetto dell'indagine. Tutti i campioni sono stati successivamente conservati a -80°C . Sul web, http://www.libyandates.com/variet%C3%A0_datteri_progetto_palma_dattero_libia.html), sono presenti delle schede per ogni varietà campionata che contengono immagini relative alla varietà in questione, datteri e infiorescenze comprese, ed altre informazioni sia di carattere agronomico che pomologico, redatte dalla sezione di Coltivazioni arboree del Dipartimento di Scienze delle Produzioni Vegetali, del Suolo e dell'Ambiente Agroforestale (DI.P.S.A.) dell'Università di Firenze.

Tabella 3.1: Numero di individui analizzati nelle diverse località dell'oasi di Al-Jufra.

Varietà	Sokna	Hun	Waddan	Zellah	Al Fugha	Al Jufrah
Abel	6	7	10	0	0	23
Bamour	1	0	7	0	0	8
Berni	5	7	9	0	0	21
Bestian	7	10	16	3	4	33
Deglet	10	7	7	1	0	25
Halima	6	3	6	0	0	15
Hamria	8	9	7	5	5	34
Kathari	12	10	6	6	0	34
Noyat Meka	0	1	8	0	0	9
Omglaiab	1	4	6	0	0	11
Saiedi	4	7	6	4	0	21
Saila	3	8	0	0	0	11
Sokeri	8	2	9	0	0	19
Tagiat	7	10	11	6	5	39
Talis	8	6	6	5	0	25
Tameg	6	9	7	0	0	22
Tasferit	4	4	4	0	0	12
Zebur	6	6	2	0	0	14
Maschi	21	20	22	0	0	63
Totale	123	131	142	30	14	440

Figura 3.2: Distribuzione nelle varie località delle varietà selezionate.



3.2 Estrazione del DNA.

Il materiale vegetale è stato polverizzato tramite un omogenizzatore (bead-mill homogenizer TissueLyser, Qiagen, Italia). Il materiale polverizzato è stato sottoposto ad estrazione del DNA totale usando sia il kit DNeasy Plant Maxi/Mini (Qiagen) che il kit E-Z 96 Plant (Omega), in accordo con il protocollo fornito dalla casa produttrice. Successivamente è stata valutata la qualità e la concentrazione del DNA purificato sia attraverso analisi spettrofotometriche a 260 nm e 280 nm che con un gel elettroforetico all'1% di agarosio.

3.3 Amplificazione e genotipizzazione dei loci SSRs nucleari.

Abbiamo valutato un totale di 18 coppie di primer (tabella 3.2) che amplificano loci SSRs nucleari specifici della palma da dattero, scelti sulla base del loro indice polimorfico (PIC) e numero totale di alleli, tra un gruppo di marcatori microsatelliti sviluppati da Billotte (2004) e Akkak (2009).

La reazione di amplificazione è stata eseguita su un volume totale di 14 μ L contenente: 20 ng di DNA genomico, 1X di buffer (Promega Corp. Madison, USA), 0,2 mM di dNTP (Promega), 0,05 U di Taq polimerasi (Promega), 0,07 μ M di primer forward, 0,02 μ M di primer reverse e 0,2 μ M di primer M13 marcato con un fluorocromo. Tutti i primer forward possiedono in aggiunta una sequenza universale M13 di 18 basi (5'-TGTA AACGACGGCCAGT-3') il cui compito è quello di fungere da ancora per una molecola fluorescente (6'-FAM, NED o VIC) che verrà incorporata durante i cicli PCR nelle sequenze microsatellite da analizzare rendendole quindi visibili al sequenziatore durante l'elettroforesi capillare permettendo anche di analizzare più marcatori contemporaneamente (Schuelke, 2000).

La PCR è stata realizzata in un termociclatore dell'Applied Biosystem (AB System) con le condizioni termiche elencate in tabella 3.3. E' stato eseguito, per ogni primer testato, anche un controllo negativo, escludendo dalla mix di reazione il DNA, con il fine di valutare la presenza di dimeri dei primer.

Per valutare la presenza della banda di dimensioni attese, abbiamo caricato i prodotti di amplificazione su un gel di agarosio all'1,5%.

Tabella 3.2: Primer considerati per la genotipizzazione del materiale fenicolo libico.

Ref.	Locus	Motivo SSR ripetuto	Primer (5' - 3')
Akkak et al. (2009)	PDCAT1	(TC) 21	F:CTGAAATCTCTGTTCAAATCC R:AGTTTGGATCTATTTGTGAGTATTTTCTTT
	PDCAT2	composito	F:GGCCTTCTCTCCCTAATGGG R:AGTTTCTTGCCCTTCTCTTC
	PDCAT6	(CA) 14 (GA) 23	F:AATCAGGGAAACCACAGCCA R:GTTTAAAGCCTTCTCAAGATAGCCTCAG
	PDCAT8	(TC) 16	F:GCTTAAAGTGGTTAGTTGCCAA R:GTTTGGCAGAGTATTGAAAAGTTGA
	PDCAT11	(TC) 7 (TC) 20	F:TTAGTAGACTCCCCACCGTCC R:TGTTTCATGGTGCTGGAGAATGAA
	PDCAT13	composito	F:TGTTGCCATTACATGTCTGCG R:TTTGGACTAGTCCCTCCCTCC
	PDCAT14	(TC) 19 (TC) 16	F:TGCTGCAATCTAGGTCACGAG R:TTTACCCCTCGGCCAAATGTAA
	PDCAT17	(GA) 21	F:CAGCGAGGGTGGGCCCTCGTT R:TCTCCATCTCCCTTTTCTCTGCTACTC
	PDCAT18	composito	F:CCTAAACCTGAATGAATCAAG R:CAACTAACATAAGGACAGTGCTATGTGATTG
Billotte et al., (2004)	mPdCIR010	(GA) 22	F:ACCCCGGACGTGAGGTG R:CGTCGATCTCCTCTTGTCTC
	mPdCIR015	(GA) 15	F:AGCTGGCTCCTCCCTCTTA R:GCTCGGTTGGACTTGTCT
	mPdCIR025	(GA) 22	F:GCACGAGAAGGCTTATAGT R:CCCCTCATTAGGATTCTAC
	mPdCIR032	(GA) 19	F:CAAATCTTTGCCGTGAG R:GGTGTGGAGTAATCATGTAGTAG
	mPdCIR48	(GA) 32	F:CGAGACCTACCTCAACAAA R:CCACCAACCAATCAAAACAC
	mPdCIR070	(GA) 17	F:CAAGACCCAAGGCTAAC R:GGAGGTGGCTTGTAGTAT
	mPdCIR078	(GA) 13	F:TGGATTTCCATTGTGAG R:CCCGAAGAGACGCTATT
	mPdCIR085	(GA) 29	F:GAGAGAGGGTGGTGTATT R:TTCATCCAGAACACAGTA
	mPdCIR093	(GA) 16	F:CCATTTATCATTCCTCTCTTG R:CTTGGTAGCTGCGTTTCTTG

Tabella 3.3: Profilo termico PCR dei primer microsatelliti mPdCIR (Billotte) e PDCAT (Akkak).

Primer	denaturazione		annealing		estensione		
	T(°C)	t (min)	cicli	T(°C)	t (sec)	T(°C)	t (min)
Billotte	95	1	35	94	30	72	8
				52	60		
				72	120		
Akkak	95	8	30-33	95	30	72	45
				55	45		
				72	60		

La genotipizzazione è stata ottenuta risolvendo i prodotti di PCR sia su un sequenziatore MegaBace 1000 (GE Healthcare, USA) che con un ABI Prism 3130 (Applied Biosystem, Foster City, CA, USA) grazie a uno standard di peso molecolare noto. L'assegnazione delle dimensioni degli alleli è stata ottenuta, sulla base del tipo di sequenziatore, utilizzando il software Fragment profiler ver.2.1 (Amersham Biosciences) e/o Genescan ver. 3.7 (Applied Biosystem, Foster City CA, USA).

3.4 Genotipizzazione del locus minisatellite plastidiale.

E' stato caratterizzato un locus minisatellite plastidiale in 84 piante appartenenti alle 18 varietà, le piante maschili non sono state inserite in queste analisi. Abbiamo incluso nell'analisi almeno due piante per varietà. Il marcatore minisatellite plastidiale è stato identificato all'interno della regione intergenica (IGS) trnG-trnfM grazie a una coppia di primer universali sviluppata da Demesure e collaboratori (1995).

Le sequenze dei primer utilizzati per caratterizzare il minisatellite nella palma sono state sviluppate da Henderson e collaboratori (2006):

primer forward → psbZ-IGS- 5'-GGTACMTCATTATGGATTGG- 3'

primer reverse → trnfM-IGS-5'-GCCGAGTAGAGCAGTTTGGT-3'.

La reazione di amplificazione è stata eseguita in un volume finale di 14 uL contenente: 20 ng di DNA, 1X PCR buffer (Promega, USA), 0.02 mM di dNTP (Promega), 0.2 uM di primer forward, 0.2 uM di primer reverse e 0.05 U di Taq DNA polimerasi. La PCR è stata fatta utilizzando un termociclature dell'Applied Biosystem (AB System, Germania) con il seguente gradiente termico:

94°C per 4 minuti, seguiti da 30 cicli a 94°C di 30 secondi, 60°C per 30 secondi e 72°C per 45 secondi seguiti da un estensione finale a 72°C per 8 minuti. Gli amplificati sono stati controllati su un gel di agarosio all'1.5% per valutare sia la dimensione attesa di 700 bp che per controllarne la qualità. La purificazione dei prodotti di PCR è stata eseguita con il sistema MultiScreen96 PCR plates e MultiScreen vacuum manifold, (Millipore) secondo il protocollo fornito dalla casa produttrice. La reazione di sequenza, fatta con entrambi i primer, è stata eseguita utilizzando il BigDye 3.1 terminator kit (Applied Biosystems). La purificazione dei prodotti di sequenza è stata fatta utilizzando la resina Sephadex-G50 (GE Healthcare, Life Sciences) con piastre multiscreen (Millipore), secondo il protocollo

fornito dalla casa produttrice e sono state sequenziate sull'ABI prism 3130 (Applied Biosystems, Foster City, CA, USA). La qualità del sequenziamento è stata valutata con il software Sequencing analysis 5.3.1 (Applied Biosystems). Le sequenze forward e reverse sono state assemblate con l'ausilio del software CodonCode Aligner (CodonCode Corporation, Dedham, MA, USA), allineate con ClustalX (Thompson et al., 1997) ed editate manualmente con il software opensource Bioedit 7.0.0 (Ibis Biosciences, Carlsbad, CA, USA).

3.5 Analisi dei dati.

La caratterizzazione genetica è stata effettuata determinando la diversità allelica (numero totale di alleli e loro frequenze all'interno della popolazione), l'eterozigosi osservata e attesa (Hobs e Hexp) (Nei 1978) utilizzando i software open-source GenAlEx 6 (Peakall & Smouse 2006), CERVUS 3.0 (Kalinowski et al., 2007) e Genetix 4.04 (Belkhir et al., 2000). Con il fine di computare una classificazione gerarchica delle varietà, abbiamo calcolato le distanze genotipiche codominanti e utilizzato la matrice delle distanze genetiche per costruire un dendrogramma attraverso il software MEGA ver.5 (Tamura et al., 2011) che implementa l'algoritmo UPGMA (Unweighted Pair Group Method with Arithmetic mean). Per avere un'idea della struttura genetica delle varietà, abbiamo utilizzato STRUCTURE 2.3.4 (Pritchard et al., 2000; Falush et al., 2003) un software di clustering bayesiano che considera l'equilibrio di Hardy-Weinberg e il linkage disequilibrium. E' stato utilizzato un modello misto (admixture model), nel quale la frazione di similarità di ogni ipotetico cluster è stimata per ogni individuo, con le frequenze alleliche correlate. Si sono implementate quattro corse indipendenti per ogni K (cluster), con il valore di K che va da 1 a 25, un burn-in-period di 50000 steps seguiti da 100000 simulazioni MCMC (Markov Chain Monte Carlo). Abbiamo utilizzato un valore di K così elevato per valutare le possibili sottostrutture genetiche che si possono presentare soprattutto nel gruppo dei maschi o in quelle varietà con una variabilità interna elevata, tipo la Sokeri che a priori conosciamo essere propagata per seme. I risultati delle simulazioni sono serviti come input per il programma Structure Harvester (Earl & vonHoldt 2012) che verifica la consistenza delle simulazioni tra le differenti repliche e identifica

il numero di cluster (K) che meglio si adatta con i nostri dati (Evanno et al., 2005).

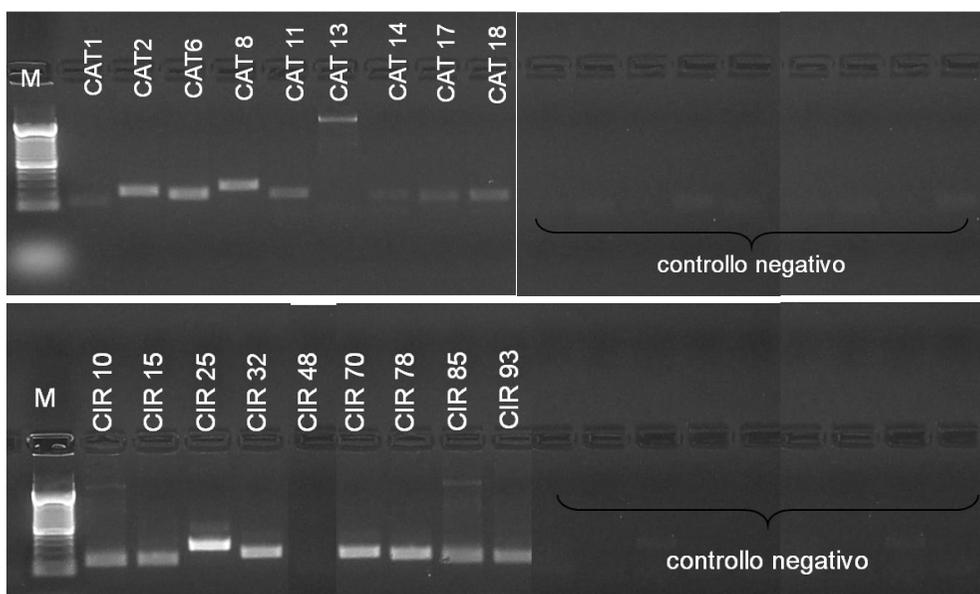
Sono state eseguite delle analisi di assegnazione varietale sulle piante maschili attraverso test di paternità che si basano sulla massima verosimiglianza per stimare l'appartenenza a una data cultivar. L'attribuzione delle piante maschili è stato eseguito confrontando i genotipi maschili con quelli delle cultivar attraverso il software CERVUS 3.0. La significatività dei valori di LOD è stata trovata attraverso simulazioni, utilizzando come parametri: 10000 ripetizioni, 0.01 come quota massima di loci con errore e 61 profili individuali come probabili cultivar candidate per ogni individuo maschile. Abbiamo scelto due livelli di confidenza, uno stringente del 95% e uno dell'80%, come suggerito da Marshall (1998). Il valore di LOD è stato ottenuto prendendo il logaritmo naturale del rapporto totale di verosimiglianza. Valori positivi di LOD stanno a significare che la varietà candidata ha elevata probabilità di essere quella vera, mentre valori negativi di LOD si ottengono quando la varietà candidata e la pianta maschile non corrispondono per uno o più loci. E' stato eseguito un test statistico sulla base delle simulazioni di appartenenza che ha permesso assegnare le piante maschili alle varietà con un buon livello di attendibilità.

4. RISULTATI

4.1 Analisi della diversità genetica.

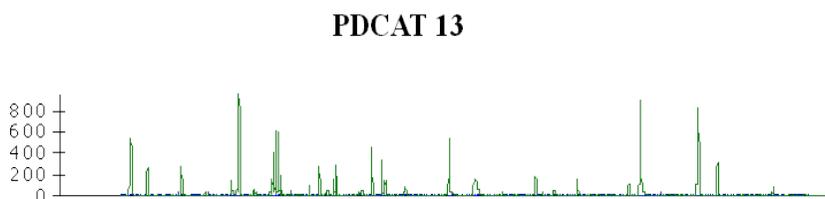
In un'analisi preliminare 18 coppie di primer (Figura 4.1) sono state valutate per la possibilità di generare un pattern SSRs chiaro, del materiale in esame. Successivamente in base ai risultati ottenuti abbiamo selezionato 16 marcatori microsatelliti migliori scartando due coppie di primer. Il locus mPdCIR48 che non ha prodotto nessuna amplificazione nei nostri campioni, come riportato anche in un lavoro di caratterizzazione delle varietà tunisine (Zehdi et al., 2004), e il locus PDCAT13, che ha dato problemi di segnali multipli in fase di genotipizzazione (Figura 4.2)

Figura 4.1: Controllo su gel di agarosio delle dimensioni attese dei 18 microsatelliti testati nelle analisi preliminari.



Note: M rappresenta il marker di peso molecolare noto (1kbase).

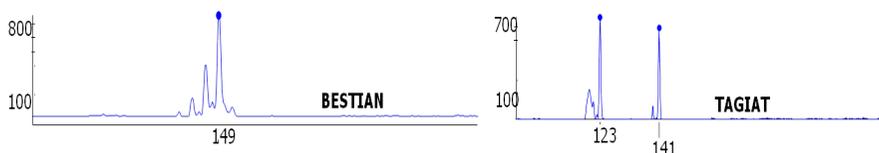
Figura 4.2: Elettroferogramma indecifrabile prodotto dal marcatore PDCAT13 ottenuto in fase di genotipizzazione.



Ottenuti i profili allelici delle varietà, sono stati rilevati due errori in fase di campionamento nel riconoscimento varietale. Una varietà riconosciuta dagli agronomi locali come Abel (H3F-H-04) e un'altra come Bamour (WHB-C-57), si sono rivelate essere rispettivamente una Zebur e una Omgluib. Quindi in fase di analisi dei dati questi due individui sono stati trattati in base alla loro varietà reale.

I profili SSRs hanno mostrato differenti alleli per locus nel campione analizzato, con la condizione omozigote ed eterozigote chiaramente identificabile a livello individuale (Figura 4.3).

Figura 4.3: Esempio di condizione omozigote e eterozigote mostrata dal marcatore PDCAT18 su due varietà. I numeri in basso rappresentano gli alleli registrati dal sequenziatore.



La diversità genetica, per ogni locus microsatellite, valutata come numero di alleli e genotipi è presentata in tabella 3.1. Il numero di alleli totale registrato è stato di 130, con una media di 8,22 alleli per locus. Il numero di alleli per locus varia da 4 per il locus PDCAT1 a 14 per il locus mPdCIR78; i valori di Hexp vanno da 0,43 (PDCAT14) a 0,87 (mPdCIR85) indicando che il germoplasma fenicolo libico è caratterizzato da un elevato grado di diversità genetica. Considerando solamente le piante maschili, abbiamo trovato un totale di 120 alleli, con una media di 7,5 alleli per locus e un numero di 20 alleli esclusivi della popolazione di piante maschili (Tabella 4.2). All'interno delle cultivar, il numero di alleli totale registrato è stato di 110, con una

media di 6,88 alleli per locus. Inoltre 28 alleli su 110 sono risultati fissati all'interno di una determinata varietà. Un esempio interessante è rappresentato dal locus PDCAT1 dove tutti gli alleli, anche in cultivar differenti, sono fissati (Tabella 4.1).

La distribuzione delle frequenze alleliche dei 16 loci SSRs nelle cinque località di Al-Jufrah (Sokna, Hon, Waddan, Zella e Al-Fugha) e nelle palme maschili campionate è presentata in figura 4.4. All'interno delle località il numero e la frequenza degli alleli varia in funzione del tipo di cultivar presente nella località considerata (Figura 2.3). Un esempio è rappresentato dal locus mPdCIR10, il quale esibisce 6 alleli, tre dei quali presenti solo nelle località principali.

Tabella 4.1: Descrizione della variabilità allelica in 16 marcatori microsatellite stimata nelle 18 varietà.

Locus SSR	Allelic range (bp)	Na	n° di genotipi	Hobs	Hexp	PIC	n° alleli fissati	Allele fissato	Cultivar con allele fissato
mPdCIR10	138-176	6	13	0,44	0,48	0,45	1	154	Abel Bemì Deglet Halima Omglab Saiedi Tagiat Tameg
mPdCIR15	142-157	6	15	0,85	0,77	0,73	2	142 157	Tagiat Zebur
mPdCIR25	219-257	6	17	0,87	0,75	0,71	1	249	Bemì
mPdCIR32	306-321	6	13	0,72	0,68	0,61	2	316 309	Saiedi Zebur Talis
mPdCIR70	205-227	9	32	0,90	0,83	0,81	1	213	Deglet
mPdCIR78	126-173	14	36	0,86	0,86	0,84	2	136 153	Tagiat Saila
mPdCIR85	175-199	12	39	0,83	0,87	0,85	2	175 181	Omglab Halima
mPdCIR93	181-197	7	17	0,77	0,77	0,74	1	197	Omglab
PDCAT1	103-123	4	10	0,23	0,63	0,59	4	103 105 119 123	Tameg Zebur Bestian Saila Abel Deglet Halima Saiedi Talis
PDCAT2	186-209	8	20	0,83	0,79	0,76	2	186 206	Zebur Bestian
PDCAT6	142-172	8	17	0,80	0,71	0,66	2	150 158	Zebur Omglab
PDCAT8	222-258	10	14	0,75	0,69	0,64	2	246 252	Saila Talis
PDCAT11	154-177	6	20	0,75	0,79	0,75	2	154 159	Zebur Saila
PDCAT14	141-163	9	20	0,43	0,64	0,60	1	155	Deglet Halima Omglab Tagiat Talis Tameg
PDCAT17	131-157	8	14	0,48	0,64	0,59	3	141 147 157	Saila Abel Deglet Kathari Noyat Meka Omglab Talis
PDCAT18	123-149	11	29	0,86	0,77	0,74	0	0	
Totale		130					28		

Na:numero di alleli; Hobs:eterozigosi osservata; Hexp:eterozigosi attesa sotto l'equilibrio di HW; PIC:polymorphic information content

Tabella 4.2: Lista di alleli trovati esclusivamente nelle piante maschili.

Locus SSRs	Alleli privati maschili
mPdCIR10	-
mPdCIR15	149
mPdCIR32	306
mPdCIR78	126; 132; 173
mPdCIR85	177; 179; 191; 199
mPdCIR93	-
PDCAT2	203
PDCAT6	152
PDCAT8	234; 236; 240; 248
PDCAT14	-
PDCAT17	149; 153
PDCAT18	125; 129; 131

Dalla figura 4.4 si possono apprezzare le frequenze degli alleli nelle località e nelle piante maschili. Un esempio è rappresentato dal locus mPdCIR85, che presenta un totale di 12 alleli, 4 esclusivi degli individui maschili, tra cui due di questi (allele 177 e allele 199) con una frequenza elevata e gli altri due, (allele 179 e allele 191) con una frequenza inferiore allo 0,5. La ricchezza allelica delle piante maschili valutata tra le tre località di campionamento non varia in maniera significativa, ma è omogenea tra le località.

Il numero medio di alleli varia da una cultivar ad un'altra, in funzione soprattutto dell'identità clonale delle varietà (Tabella 4.3). I risultati mettono in evidenza la differente struttura genetica delle cultivar.

Figura 4.4: Distribuzione delle frequenze alleliche dei microsattelliti nelle 5 località di Al-Jufrah e nel gruppo delle palme maschili (“male”).

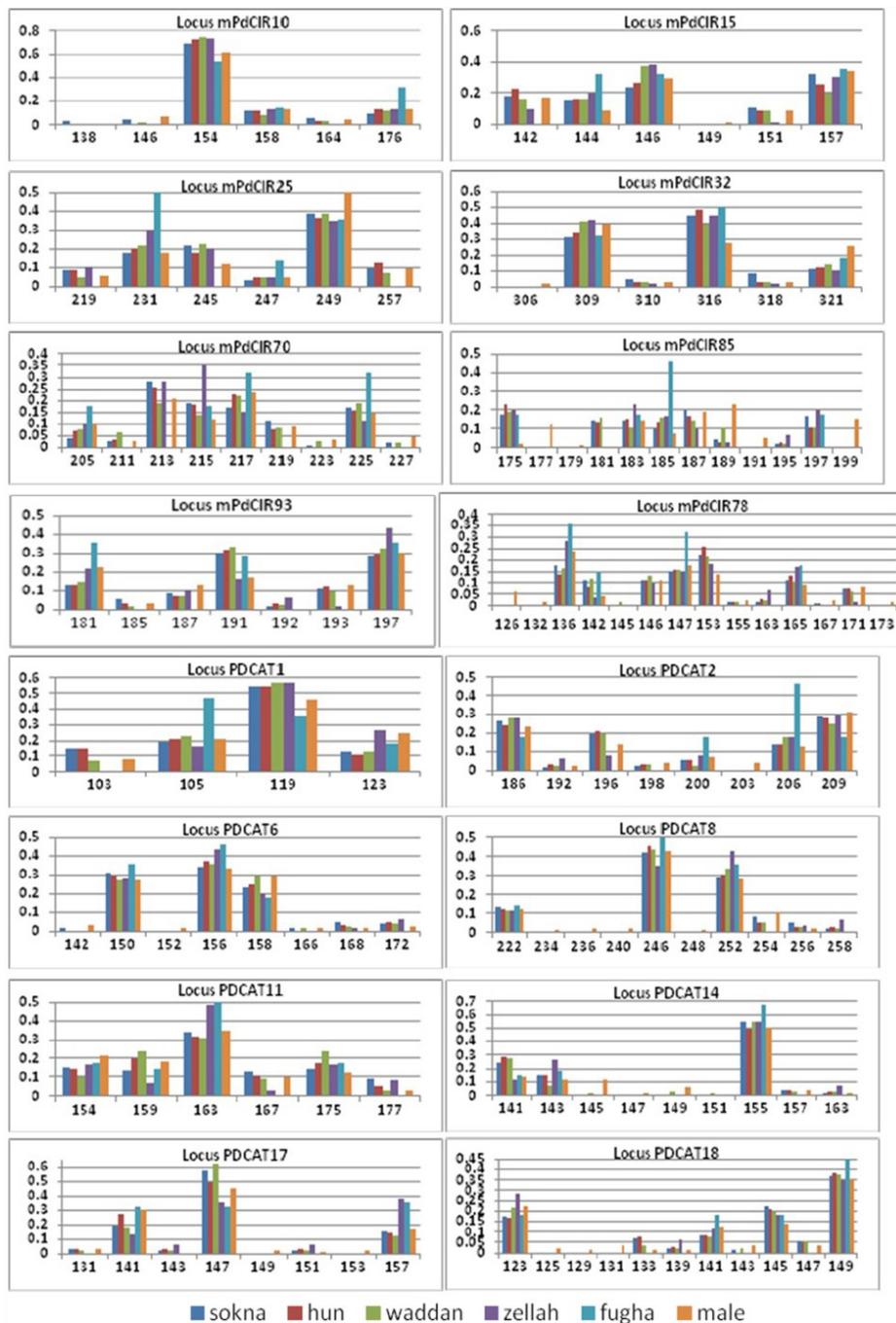


Tabella 4.3: Indici di diversità genetica calcolati per le 18 cultivars e per il gruppo di piante maschili.

Cultivar	N	Na	Ne	Hobs	Hexp	UHexp	F
Abel	23	1,937	1,857	0,747	0,379	0,387	-0,974
Bamour	7	2,813	2,121	0,653	0,491	0,530	-0,268
Berni	21	2,625	1,733	0,622	0,348	0,356	-0,567
Bestian	33	2,750	1,684	0,629	0,339	0,344	-0,557
Deglet	25	1,688	1,630	0,625	0,314	0,320	-0,993
Halima	15	1,750	1,750	0,750	0,375	0,388	-1,000
Hamria	34	3,313	2,049	0,915	0,497	0,504	-0,799
Kathari	34	2,500	1,861	0,807	0,425	0,432	-0,747
Noyat Meka	9	2,563	1,755	0,576	0,365	0,387	-0,368
Omglab	12	1,563	1,563	0,563	0,281	0,293	-1,000
Saiedi	21	1,813	1,813	0,813	0,406	0,416	-1,000
Saila	11	1,625	1,625	0,625	0,313	0,327	-1,000
Sokeri	19	5,125	3,630	0,708	0,656	0,674	-0,007
Tagiat	39	1,813	1,813	0,813	0,406	0,412	-1,000
Talis	25	1,625	1,625	0,625	0,313	0,319	-1,000
Tameg	22	2,750	1,833	0,685	0,380	0,389	-0,670
Tasferit	12	3,250	1,924	0,630	0,420	0,438	-0,240
Zebur	15	1,625	1,625	0,625	0,313	0,323	-1,000
Maschi	63	7,500	4,268	0,695	0,744	0,750	0,069

N: numero di individui; **Na:** numero di alleli; **Ne:** numero di alleli effettivi; **Hobs:** eterozigosi osservata; **Hexp:** eterozigosi attesa sotto l'equilibrio di HW, **UHexp:** eterozigosi attesa non distorta; **F:** indice di fissazione.

Tutte le cultivar sono caratterizzate da un valore negativo dell'indice di fissazione (F) a causa di un eccesso di eterozigoti rispetto all'equilibrio di Hardy-Weinberg (HW), anche se a differenti livelli. In particolare le cultivars Abel, Deglet, Halima, Omglab, Saiedi, Saila, Tagiat, Talis e Zebur, presentano un indice di fissazione prossimo o uguale a -1, che indica una forte pressione eterotica alla base del miglioramento clonale in queste varietà. Dall'altra parte un valore di F vicino o superiore allo zero è atteso in caso di incroci casuali, come osservato nella cultivar Sokeri che è tradizionalmente propagata via seme.

Una rappresentazione grafica ottenuta mediante il “package Structure” dei genotipi delle cultivar e delle piante maschili è mostrato in figura 4.5.

Ogni istogramma rappresenta una pianta. Tutte le cultivars che presentano una colorazione omogenea sono composte da piante che condividono lo stesso pattern allelico, non c'è variazione genetica all'interno della cultivar come atteso in base alla propagazione clonale. Le cultivar omogenee sono Abel, Deglet, Halima, Omglaiib, Saiedi, Saila, Tagiat e Zebur, dati che sono in accordo con gli indici di fissazione riportati in tabella 4.3.

Questa rappresentazione grafica, oltre ad evidenziare le piante che non condividono completamente il profilo varietale, vedi ad esempio l'individuo appartenente alla cv Tasferit che condivide "blocchi" della cv Bestian e Sokari, consente anche di fare altre considerazioni se, ad esempio si tiene conto anche della località in cui ciascuna pianta è stata campionata. E' il caso, ad esempio, della varietà Bestian, con 31 piante che condividono lo stesso profilo allelico(96% di identità clonale), ma con 2 individui campionati in un'azienda della località di Zella che presentano una differente composizione allelica per alcuni loci marcatori. Similmente, la varietà Sokari presenta un'elevata variabilità sia tra le località che al loro interno (Figura 4.6). Questa cultivar presenta due genotipi principali di sicura provenienza clonale, localizzati in tutte le aziende della località di Sokna e un gruppo di piante eterogenee a causa della propagazione per seme fatta abitualmente nella località di Wadan.

Figura 4.5: Analisi di clustering bayesiana condotta su tutti gli individui dell'oasi di Al-Jufrah.

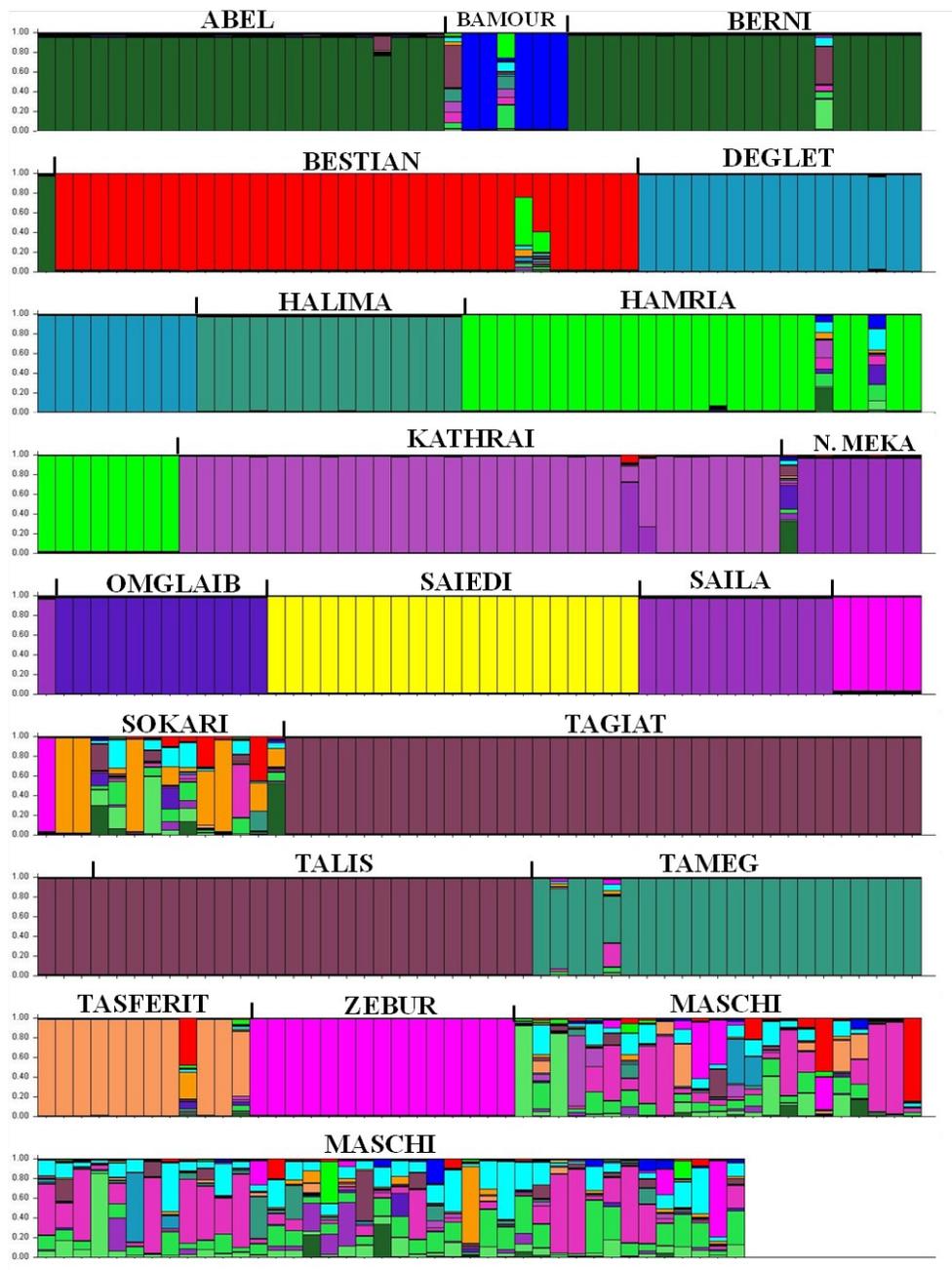
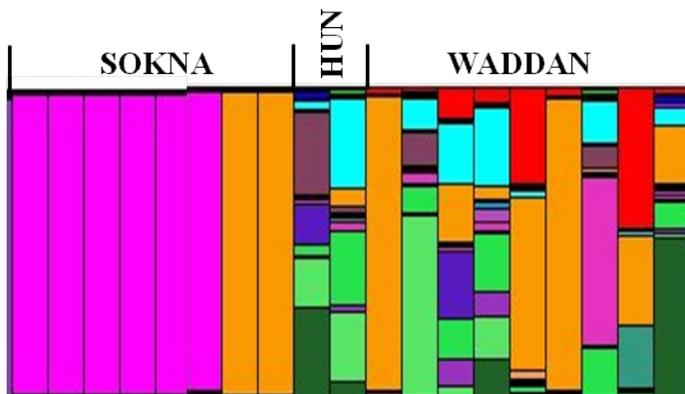


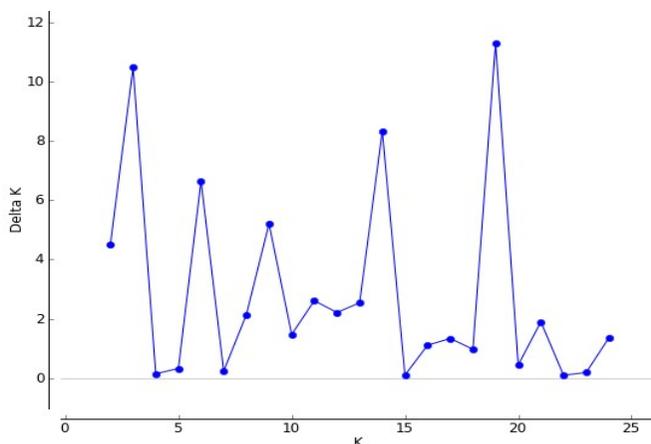
Figura 4.6: Struttura genetica della cv Sokari nelle tre località di campionamento.



Il gruppo delle piante maschili presenta molta variabilità con un elevato grado di mescolamento e si notano dei “blocchi” specifici fissati in tutta la popolazione.

Il numero stimato di cluster (K) è uguale a 19, ed è in accordo con il numero totale di gruppi analizzati, cioè le 18 cultivar e il gruppo di piante maschili (Figura 4.7).

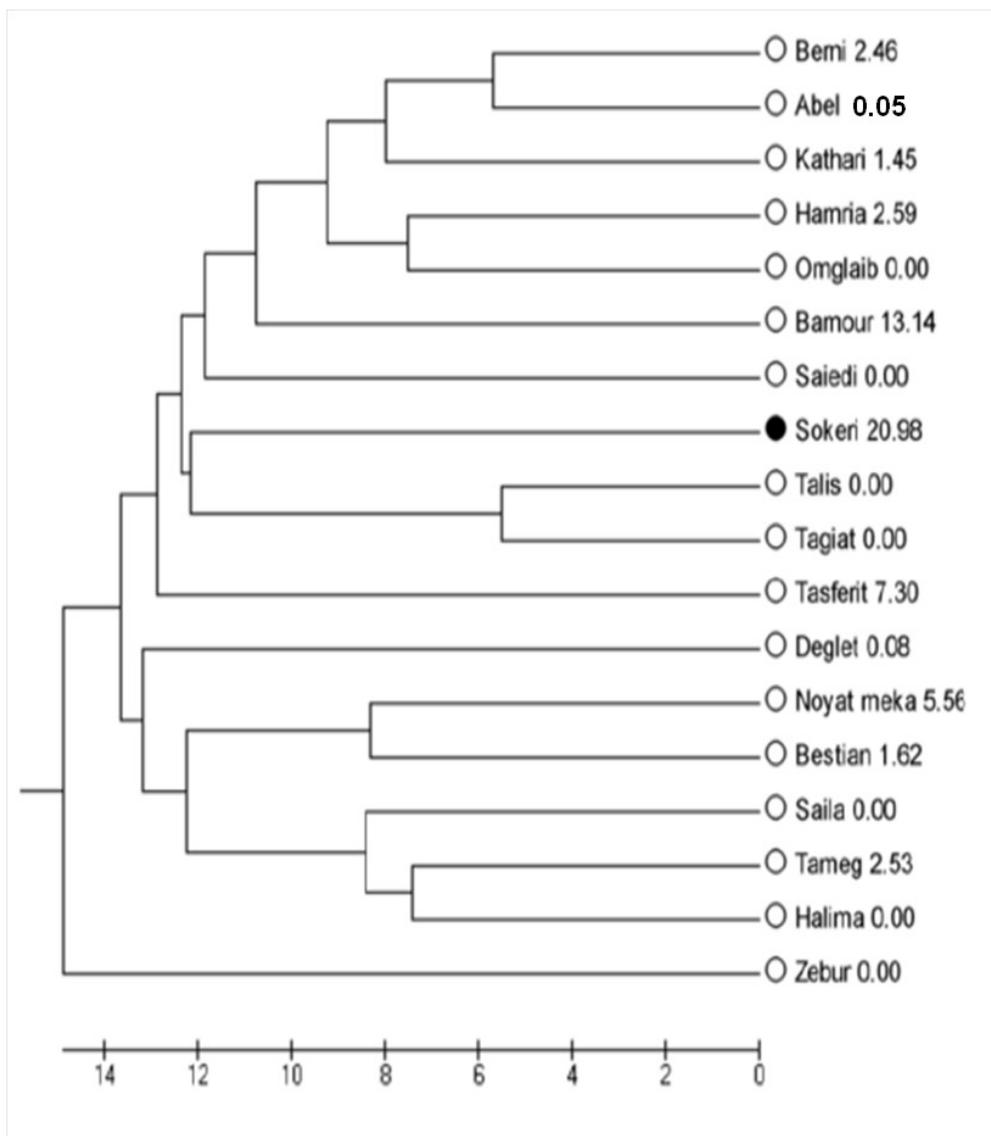
Figura 4.7: Stima del numero più probabile di K (cluster). Il picco più alto è quello relativo a K=19.



Il software utilizzato evidenzia con lo stesso colore cultivar simili sulla base del numero di alleli condivisi (Abel e Berni, Halima e Tameg, Noyat Meka e Saila, Tagiat e Talis). I risultati di similarità ottenuti con questo metodo di

inferenza bayesiana, sono concordanti con quelli ottenuti utilizzando altri algoritmi, come il metodo UPGMA (Figura 4.8).

Figura 4.8: Dendrogramma UPGMA delle 18 cultivar femminili utilizzando le distanze genotipiche codominanti.



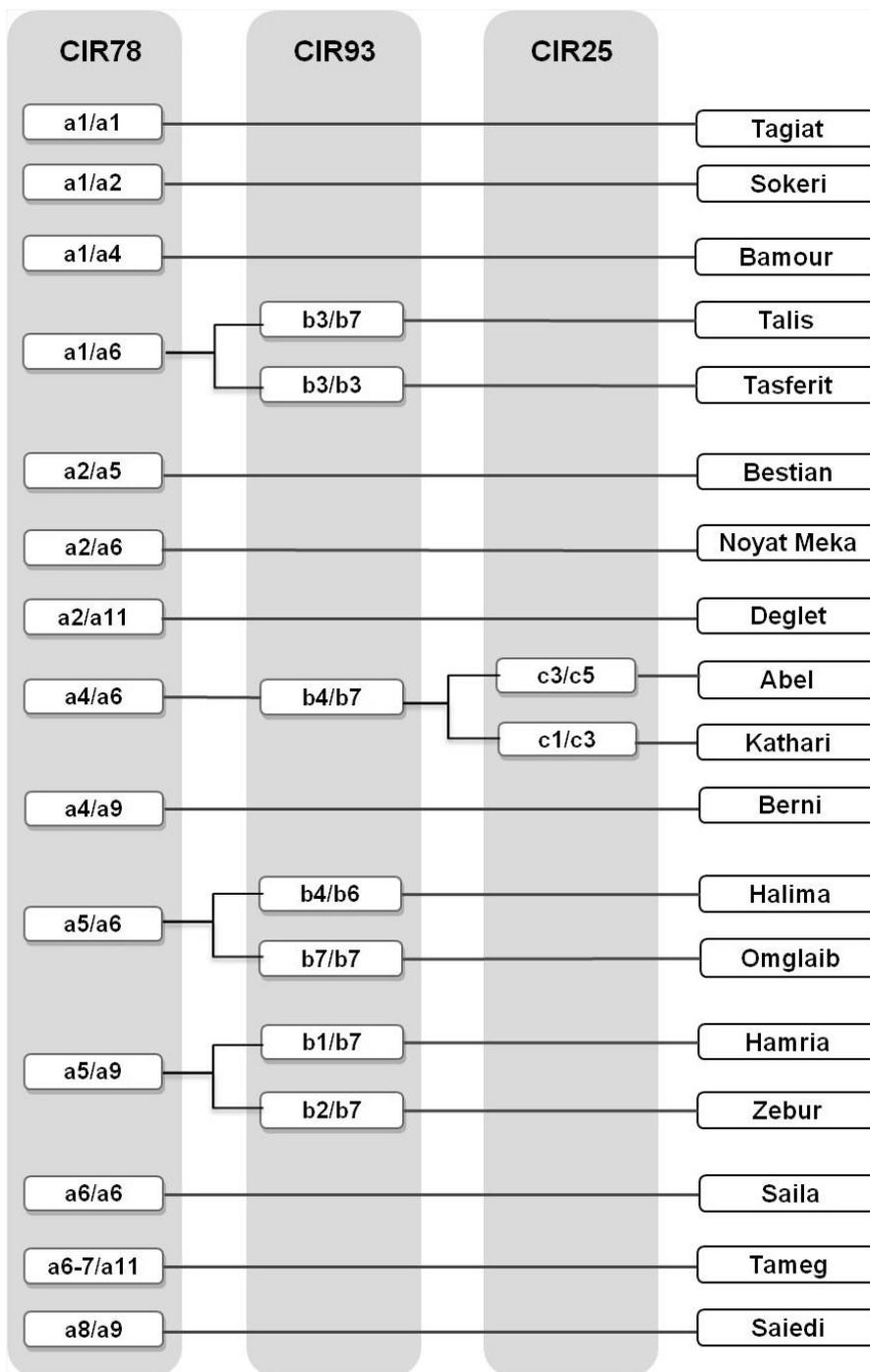
Il dendrogramma UPGMA è stato costruito con le distanze genotipiche codominanti utilizzando i 110 alleli SSR ritrovati nelle cultivar. La topologia del cluster mette in evidenza la diversità genetica esistente tra le cultivar che permette una facile distinzione tra queste. Le distanze genotipiche

codominanti permettono di stimare una similarità media all'interno di ogni varietà che va da 0 a 20,98, ed è presentata nel cluster della figura 4.7 con un indice. Abel, Deglet, Halima, Omglab, Saiedi, Saila, Tagiat e Zebur mostrano un valore prossimo allo zero, indicando che non c'è differenza genetica all'interno delle cultivar, in accordo sia con le analisi bayesiane che con gli indici di fissazione (F). Il valore più elevato è stato osservato per la cv Sokari (20,98), ed è un valore atteso sulla base del modo di propagazione per seme che caratterizza la cultivar.

4.2 Chiave identificativa delle 18 cultivar.

Per ciascuna delle 18 cultivar è stato rilevato il genotipo mostrato dai loci SSR mPdCIR78, mPdCIR93 e mPdCIR25. Abbiamo identificato 23 alleli in questi loci: 10 alleli (etichettati da a1 a a11) per il marcatore mPdCIR78, 7 alleli (etichettati da b1 a b7) per il marcatore mPdCIR93 e 6 alleli (etichettati da c1 a c6) per il marcatore mPdCIR25. Sulla base delle combinazioni alleliche siamo stati in grado di stabilire una chiave identificativa specifica per le 18 cultivar considerate (Figura 4.9). Gli individui della cultivar Sokari inseriti in questa analisi sono solo quelli appartenenti alla località di Sokna. Tutte le cultivar esibiscono un profilo differente attraverso i loci SSR considerati, confermando la capacità di questi marcatori di identificare i genotipi.

Figura 4.9: Chiave identificativa delle cultivar.



4.3 Assegnazione delle piante maschili ad una varietà.

Considerata l'elevata efficienza mostrata dai marcatori microsatelliti nella determinazione del fingerprinting genotipico delle cultivar, li abbiamo utilizzati per assegnare le piante maschili a una determinata cultivar.

Il metodo utilizzato ha permesso di assegnare a una data cultivar con elevata probabilità 55 piante maschili su un totale di 63. In tabella 3.4 sono riportati i risultati dell'analisi relativa a 24 piante maschili con valori di LOD positivi.

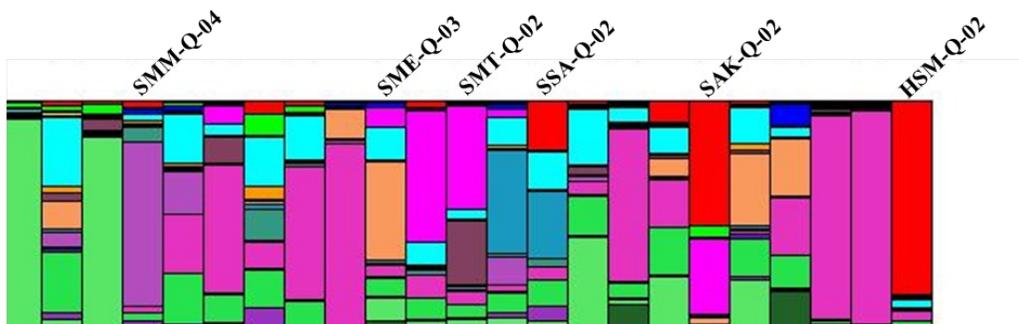
Questi dati sono in accordo con l'assegnazione ottenuta dall'analisi di clustering bayesiano, la quale a differenza dell'assegnazione basata sulla massima verosimiglianza, permette di indagare a fondo la natura ibrida di questi individui.

La figura 4.10 mostra, per un sotto-campione di piante maschili, l'assegnazione ottenuta con il metodo bayesiano. I 6 istogrammi con il codice identificativo corrispondono ai maschi evidenziati in tabella 3.4. Si può notare, ad esempio, come la pianta SMM-Q-04 sia assegnata, con oltre l'80% di corrispondenza genotipica, alla cultivar Kathari e la pianta HSM-Q-02, con più del 90% di identità, alla cultivar Bestian, sulla base delle corrispondenze varietali mostrate in figura 4.5. La pianta SSA-Q-02 condivide due "blocchi" principali: una percentuale maggiore corrispondente alla cv Deglet, che è anche quella individuata con il metodo basato sulla massima verosimiglianza, e un "blocco" minore appartenente alla cv Bestian. Quindi viene proposto un potente metodo di assegnazione, basato su entrambi i modelli.

Tabella 3.4: Assegnazione dei maschi.

Male ID	Candidate cultivar	Pair loci compared	Pair loci mismatching	Pair LOD score
SMM-Q-04	Kathari	15	0	9.47
SME-Q-03	Tasferit	16	0	0.13
SMT-Q-01	Tasferit	16	2	6.60
SMT-Q-02	Sokeri	15	1	1.26
SSA-Q-01	Deglet	15	1	9.87
SSA-Q-02	Deglet	16	2	1.77
SAK-Q-02	Bestian	16	1	6.97
SAK-Q-03	Sokeri	16	1	2.20
HSM-Q-02	Bestian	16	1	1.26
HRG-Q-01	Tagiat	16	1	3.08
H6I-Q-02	Deglet	16	2	1.05
H6I-Q-03	Tagiat	16	1	5.73
H6E-Q-03	Bamour	16	2	1.33
H3F-Q-03	Tameg	16	1	1.26
H5H-Q-02	Abel	16	1	3.45
H5H-Q-03	Hamria	16	2	2.16
H3H-Q-02	Bestian	16	1	1.12
WOE-Q-02	Abel	16	0	9.15
WBH-Q-01	Abel	16	1	3.33
WBH-Q-02	Sokeri	16	2	1.75
WFZ-Q-03	Tagiat	16	1	9.90
WHS-Q-01	Tagiat	16	1	3.80
W4B-Q-03	Hamria	16	1	1.82
WBB-B	Tasferit	16	0	1.60

Figura 4.70: Assegnazione Bayesiana di una parte di individui maschili ad una varietà. La corrispondenza tra colore e varietà è quella presentata in figura 4.5.



4.4 Analisi del minisatellite plastidiale.

I risultati dell'analisi sull'allineamento del locus minisatellite sono presentati in tabella 4.5 che riassume il numero di individui per varietà analizzata e il tipo di aplotipo presentato. La figura 4.11 evidenzia le varianti aplotipiche tra le diverse specie del genere *Phoenix*.

Tabella 4.5: Numero di individui analizzati e principale aplotipo riscontrato tra le 18 cultivar libiche.

Varietà	N	VNTR
Abel	4	3
Bamour	2	3
Berni	2	3
Bestian	8	3
Deglet	5	4
Halima	3	3
Hamria	6	3
Kathari	5	3
Noyat Meka	4	3
Omglab	2	3
Saiedi	7	4
Saila	4	3
Sokeri	6	3
Tagiat	9	3
Talis	6	3
Tameg	4	3
Tasferit	4	3
Zebur	4	3

N: numero di individui analizzati; **VNTR:** variable number tandem repeat.

L'allineamento di sequenza di questa regione effettuato da Pintaud e collaboratori (dati non pubblicati, Genbank accession da EU043466 a EU043497) in varie specie di palma (Figura 4.11) ha mostrato un polimorfismo VNTR con un motivo di 12 bp (CTAACTACTATA) ripetuto, nella dactylifera, dalle 3 alle 4 volte, il quale è parte di una ripetizione invertita, simile in struttura al minisatellite ritrovato nella famiglia delle Orchidaceae (Cozzolino et al. 2003).

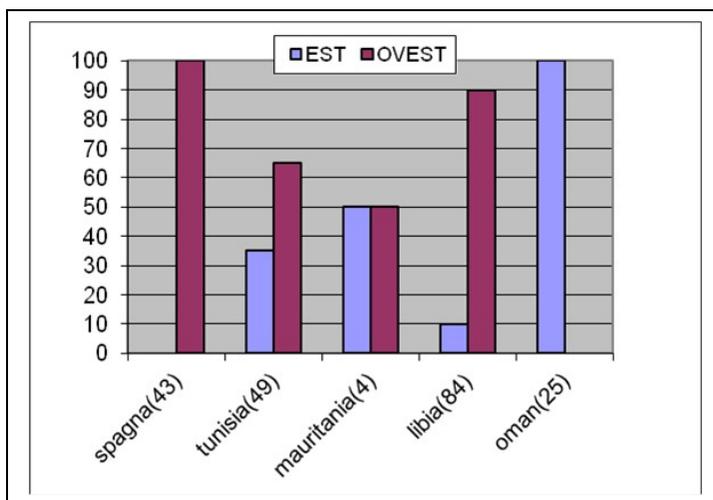
Figura 4.13: Multiallineamento del locus *psbZ-trnFM* tra alcune varietà libiche. I rettangoli colorati presenti sulla sequenza identificano il minisatellite ed è raffigurata sia la versione a 3 che a 4 ripetizioni. I rettangoli sulla destra comprendono gli individui appartenenti alle varietà che presentano l’aplotipo a 4 ripetizioni.

Abel10	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTG-----ACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.
Bern18	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTG-----ACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.
Bestian17	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTG-----ACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.
Bestian20	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTG-----ACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.
Deglet13	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTGACTACTATACTAACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.
Deglet25	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTGACTACTATACTAACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.
<< Deglet7	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTGACTACTATACTAACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.
Halima59	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTG-----ACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.
Hmria22	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTG-----ACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.
<< Nathari16 RW	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTG-----ACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.
NoyatMeka58	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTG-----ACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.
Saiedi14	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTGACTACTATACTAACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.
Saiedi16	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTGACTACTATACTAACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.
Saiedi17	CCGCTATCCGCCACGATGAAGTAATGTACTATGATAAAAAATTCGGTATGGTTGACTACTATACTAACTACTATACTAACTACTATACTAACTACTATACTAACTACTATAGT.

Se si escludono le due varietà (Deglet e Saiedi) con l’aplotipo a 4 ripetizioni le varietà autoctone libiche dell’oasi di Al-Jufrah presentano tutte l’aplotipo a 3 ripetizioni. E’ stato ipotizzato che l’aplotipo a 3 ripetizioni rispecchi un’origine ovest delle palme da dattero viceversa l’aplotipo a 4 ripetizioni un’origine orientale. Il materiale di Al-Jufrah avrebbe quindi sulla base di questa analisi un’origine occidentale.

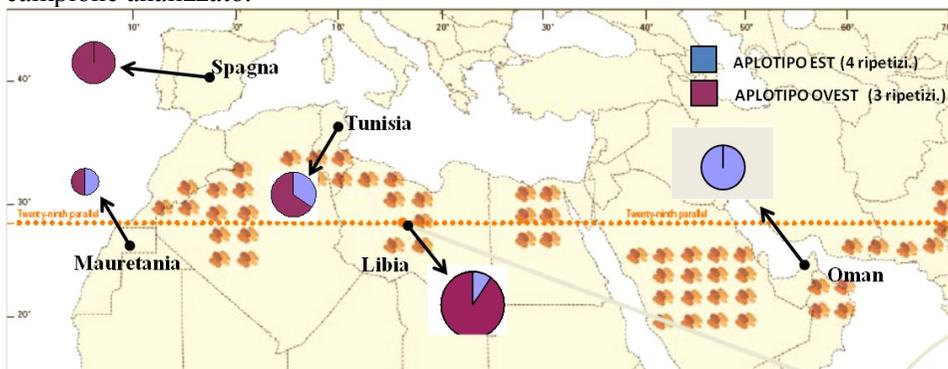
In letteratura non esistono dati sulla distribuzione degli aplotipi del minisatellite plastidiale nei paesi del Maghreb e del Medio Oriente, recentemente Ballardini e Pinteaud al convegno EUNOPS (European Network of Palm Scientists Maggio 2012) hanno presentato dati di sequenza del minisatellite in diverse provenienze di palma da dattero (Spagna, Mauritania, Tunisia e Oman) dimostrando che l’aplotipo “ovest” (3 ripetizioni) era presente in tutte le accessioni spagnole mentre quello “est” (4 ripetizioni) in tutte quelle dell’Oman (Figura 4.14).

Figura 4.14: Percentuale dei due aplotipi in accessioni di palma da dattero presenti in cinque paesi lungo il 29°parallelo. Tra parentesi è indicato il numero di individui analizzati.



La distribuzione spaziale dei due aplotipi è presentata in figura 4.15 e si può notare come l'aplotipo a 3 ripetizioni sia prevalente nelle accessioni del Nord Africa.

Figura 4.15: Distribuzione spaziale degli aplotipi est e ovest nelle accessioni provenienti da diverse aree geografiche. La dimensione delle torte è proporzionale al campione analizzato.



5. DISCUSSIONE

La palma da dattero è uno dei principali fattori dell'ecosistema oasi e della stabilità economica delle popolazioni che abitano questi ambienti estremi. Il germoplasma fenicolo libico preserva ancora oggi un'enorme ricchezza comprovata dalle oltre 400 varietà esistenti e descritte, fino ad ora, solamente utilizzando i caratteri morfologici come chiave identificativa delle cultivar. Le risorse genetiche della palma da dattero meritano di essere valutate con il fine di organizzare la loro preservazione, trasmettere questa ricchezza genetica ed anche esplorarla. Nell'oasi di Al-Jufrah, ogni palmeto è caratterizzato da una composizione varietale autoctona, dovuto principalmente alla selezione locale operata all'interno dell'oasi stessa. Sono state analizzate 18 cultivar che rappresentano i genotipi più comuni distribuiti nell'oasi attraverso l'utilizzo di 16 loci microsatelliti altamente polimorfici. E' stato inserito nelle analisi anche un sottogruppo di piante maschili di varietà ignota che rappresentano i principali impollinatori dell'oasi. Abbiamo registrato un elevato numero di alleli SSR, con una media di 8,22 alleli per locus che hanno permesso di rivelare un alto grado di variabilità genetica in queste accessioni. Questi alti livelli di polimorfismo presenti nelle cultivar sono in accordo con quelli precedentemente riportati per le cultivar mauritaniche, marocchine, algerine, tunisine, sudanesi e quatarote utilizzando marcatori SSRs (Bodian et al., 2012a, b; Akkak et al., 2009; Zehdi et al., 2004a, b; Elshibli & Korpelainen 2008; Elmeer & Mattat 2012). I risultati di ricchezza allelica sono comparabili anche con quelli riportati in altre specie longeve mediterranee come l'olivo (Bracci et al., 2009) o il fico (Achtak et al., 2009). Ogni cultivar è il risultato di una selezione empirica operata dagli agricoltori delle varie aziende dell'oasi basata sui caratteri morfologici e sulla qualità del frutto. Questi fattori giustificano la presenza di molti alleli fissati nelle cultivar, favoriti dalle procedure di propagazione clonale. Esistono diverse piante dioiche agronomicamente importanti, come il kiwi, la papaya, il luppolo e il pistacchio, in cui risulta fondamentale l'identificazione del sesso femminile negli stadi precoci di sviluppo della pianta. In questo studio abbiamo identificato 20 alleli privati in piante maschili rivelati da 9 marcatori (Tabella 4.2), che potrebbero essere utilizzati nella discriminazione precoce del sesso, come riportato in un lavoro di Elmeer & Mattat (2012).

L'analisi della struttura genetica (Figura 4.5) ha messo in evidenza che 9 cultivar su 18 hanno un'identità clonale del 100% tra le diverse località dell'oasi, in accordo con gli indici di fissazione riportati per queste cultivar (Tabella 4.3) e con gli indici di similarità calcolati sulla base delle distanze alleliche codominanti (Figura 4.8). Questi risultati evidenziano l'ottima abilità, maturata grazie all'esperienza tramandata negli anni dagli agricoltori, nelle pratiche di propagazione clonale delle cultivar; tuttavia abbiamo registrato nel nostro campione due casi di classificazione errata, che si possono verificare durante la propagazione a causa della difficoltà nella distinzione di alcune cultivar, soprattutto quelle strettamente correlate, sulla base dei descrittori pomologici. Dall'altra parte, i valori positivi dell'indice di fissazione mostrato dalla cv Sokari (Figura 4.6) e dal gruppo delle piante maschili indicano un alto livello di diversità genetica in questi individui, giustificato in relazione alle pratiche di propagazione per seme adattate in questi casi. La struttura genetica delle piante maschili (Figura 4.5) mostra, per la maggior parte degli individui, un chiaro pattern a "blocchi" dovuto all'elevato grado di ricombinazione ed eterozigosità in questo gruppo. Si può comunque notare un pattern comune fissato in tutta la popolazione maschile, che potrebbe permettere una rapida identificazione del sesso della pianta, ma tale affermazione ha bisogno di ulteriori accertamenti per essere confermata. E' stata costruita una chiave identificativa delle cultivar utilizzando tre loci microsatellite altamente polimorfici, che hanno esibito 23 alleli in totale. Tutte le cultivar sono state identificate in maniera semplice e non ambigua sulla base dei loro profili allelici. Un risultato simile è stato ottenuto precedentemente da Zhedi e collaboratori (2006) analizzando 49 accessioni tunisine con tre loci SSR. L'efficacia e la potenza discriminatoria mostrata dai microsatelliti nell'individuazione di tutte le varietà e accessioni esaminate, conferma l'utilità di questi marcatori nelle tecniche di "fingerprinting" clonale e identificazione varietale. Il trasferimento di questi marcatori può risultare di grande utilità per l'etichettatura su larga scala dei polloni e delle colture in vitro. I nostri dati evidenziano la possibilità di usare questi potenti marcatori come descrittori nella certificazione e nel controllo di origine del materiale vegetale. Nella prospettiva futura di migliorare la qualità del frutto attraverso un piano di incrocio ricorrente, le piante maschili hanno bisogno di essere identificate e assegnate a una specifica cultivar. Di conseguenza, la possibilità di attribuire a un individuo maschile di varietà

sconosciuta la probabile varietà di appartenenza è stato effettuato attraverso analisi di parentela. I risultati ottenuti permettono di assegnare 24 maschi sui 63 analizzati, a una delle 18 cultivar con un elevato livello di confidenza. Abbiamo trovato una forte correlazione tra i risultati ottenuti con le metodiche di assegnazione basati sulla massima verosimiglianza e quelli basati sull'inferenza bayesiana (Figura 4.10), mettendo in luce la possibilità di aumentare la potenza delle analisi di assegnazione attraverso l'utilizzo di entrambi questi metodi. I risultati positivi ottenuti nell'identificazione dei maschi confermano ulteriormente l'utilità dei microsatelliti per le analisi di genotipizzazione e apre nuove prospettive per il miglioramento in questa specie.

Molte ricerche avvalorano l'ipotesi di un'unica origine di domesticazione in Mesopotamia della palma da dattero (Wrigley 1995; Trifi et al., 2000; Zhedi et al., 2004). Studi più recenti effettuati su accessioni del Marocco e della Tunisia hanno invece evidenziato l'esistenza di differenti origini (Elshibli & Korpelainen 2008; Hammadi et al., 2011). In accordo con questa ultima ipotesi, abbiamo analizzato un marcatore minisatellite plastidiale di circa 700 bp proveniente da una regione localizzata nella Large Single Copy (LSC). Le due cultivar identificate come alloctone, Saiedi e Deglet, presentano entrambe l'aplotipo orientale con 4 ripetizioni (Tabella 4.5) che è in accordo con l'origine egiziana della Saiedi ma in contrasto con l'origine algerina della Deglet. I dati di distribuzione aplotipiche in alcune regioni del Nord Africa hanno mostrato che prevale quello "ovest", ovvero l'aplotipo a 3 ripetizioni (Figura 4.15) (Pinteaud & Ballardini, 2012 dati non pubblicati). Per confermare queste ipotesi, vanno comunque genotipizzati i minisatelliti plastidiali di molte altre varietà in tutto l'areale di distribuzione della *P. dactylifera*.

Alcuni risultati di questo lavoro sono stati inseriti in un paper scientifico e sottomessi per la sua pubblicazione.

Bove A, Turchi A, Bashir G, Battaglia M, Camussi A, Racchi ML (2012). Genetic characterization of Libyan date palm resources by microsatellite markers. Genetic Resource and Crop Evolution.

Ringraziamenti

Il presente lavoro di tesi si è svolto nell'ambito del Progetto "Miglioramento e valorizzazione della palma da dattero nell'oasi di Al Jufrah in Libia" coordinato dall'Istituto Agronomico per l'Oltremare (IAO) in collaborazione con l'Ente Libico per lo Sviluppo e il Miglioramento della Palma da dattero e dell'Olivo ed è stato finanziato dalla Direzione Generale per la Cooperazione allo Sviluppo del Ministero Italiano degli Affari Esteri.

CAPITOLO 2

SNP in geni candidati per valutare la diversità genetica in popolazioni naturali di *Pinus nigra*

1. INTRODUZIONE

L'importanza che le foreste ricoprono per il sostentamento della vita in generale e degli ecosistemi in particolare è molto elevata. Basti pensare che oltre il 30% della superficie terrestre è ricoperta da foreste e che queste ospitano oltre il 50% della biodiversità (FAO 2010); molte specie sono fonte di materie prime, giocano un ruolo chiave nell'interazione con svariati cicli bio-geochimici degli elementi, con ripercussioni positive sulla qualità dell'aria, dell'acqua e del suolo e sono tra i principali stocicatori di gas ad effetto serra, contribuendo in maniera sostanziale alla mitigazione degli effetti climatici (Brown et al., 1996).

Le specie forestali, a differenza della maggior parte delle piante di interesse agrario, sono poco domesticate e quindi ospitano una grande quantità di ricchezza genetica.

Le specie forestali studiate appartengono principalmente a quattro famiglie – una gimnosperma (Pinaceae) e tre angiosperme (Salicaceae, Myrtaceae e Fagaceae) – e sette generi all'interno di queste famiglie (*Pinus*, *Picea*, *Pseudotsuga*, *Populus*, *Eucalyptus*, *Quercus* e *Castanea*). Non esiste una specie modello (anche se *Populus* potrebbe esserlo) né una il cui valore economico è significativamente superiore (Neale & Kremer 2010). Per queste ragioni, poche sono le informazioni disponibili per questi organismi, in particolare relativamente alla struttura e organizzazione del genoma, anche se, con l'avvento delle nuove tecnologie di sequenziamento, questa situazione potrebbe cambiare molto rapidamente, soprattutto all'interno della famiglia delle Pinaceae.

Considerando l'importanza ecologica ed economico-sociale che queste specie hanno, diviene fondamentale stimare la loro diversità genetica per definire appropriati programmi di conservazione, in un contesto di cambiamenti ambientali repentini che minacciano anche la stessa sopravvivenza delle specie.

1.1 Introduzione sul genere *Pinus*: tassonomia e filogeografia

Il genere *Pinus* è il più rilevante all'interno della famiglia delle Pinaceae e più in generale tra le conifere sia per numero di specie (ad oggi ne sono riconosciute circa 109 (Farjon 2001)) che per il contributo nella formazione e il mantenimento degli ecosistemi. La maggior parte delle specie sono distribuite nell'emisfero boreale, raggiungendo l'equatore solo nel sud-est Asiatico. In Africa le specie sono confinate a quelle regioni che si affacciano sul bacino del Mediterraneo. Da un punto di vista della ricchezza in termini di specie sono stati identificati due centri principali, uno localizzato in Nord America e l'altro in Cina, dove sono state censite più di 50 specie. Tutte le specie di pini sono aghifoglie resinose sempreverdi, con portamento arboreo e solo raramente presentano uno status arbustivo. Alcune specie possiedono areali di distribuzione naturale molto estesi (e.g. *P. sylvestris*, *P. ponderosa*) ma la maggior parte o hanno areali di distribuzione ristretta (alcune specie confinate nell'areale Mediterraneo) o formano dei veri e propri endemismi (e.g. *P. canariensis*) (Quézel & Médail 2003). I pini sono adattati a una vasta gamma di condizioni ecologiche anche estreme, che vanno dai climi boreali (e.g. *P. pumila*), temperati (e.g. *P. thunbergii*) a quelli sub-alpini (e.g. *P. cembra*, *P. mugo*) e tropicali (e.g. *P. tropicalis*); crescono bene su suoli acidi e calcarei e si possono ritrovare su differenti tipologie di siti (rocciosi, paludosi e sabbiosi) (Richardson & Rundel 1998). Molte specie sono pioniere in quanto si insediano per prime su terreni poco ospitali o poveri di nutrienti. Alcune specie (e.g. *P. halepensis*, *P. banksiana*) manifestano una spiccata serotinia (Daskalakou & Thanos 1996), quale adattamento ecologico al fuoco che permette alle specie di rilasciare i semi dallo strobilo durante un incendio boschivo in un contesto che offre maggiori probabilità di colonizzazione (Keely & Zedler 1998). Questo adattamento oltre ad essere una forma di pionierismo, è importante dove il fuoco rappresenta un fattore ambientale determinante nel modificare le successioni degli ecosistemi naturali influenzando le dinamiche delle popolazioni.

Possono formare popolamenti puri o essere consociati, in base alla fascia climatica, con altri generi delle Pinaceae (*Picea ssp* e *Abies ssp*) e/o con altre specie di latifoglie. Sono specie longeve: metodologie dendrometriche hanno permesso di stimare che alcune, come *P. longeva* e *P. ponderosa*, possono vivere per oltre 3000 anni (Schulman 1958; Kaufmann 1996). Flanary e

Kletetschka (2005) hanno ipotizzato, confrontando le caratteristiche dei telomeri in vari gruppi di pini, che un incremento della lunghezza dei telomeri e dell'attività della telomerasi può direttamente o indirettamente contribuire ad incrementare la durata della vita in quelle specie con elevata longevità (cicli vitali che vanno da 2000 a 5000 anni) rispetto a quelle specie a media (cicli vitali che vanno da 400 a 500 anni) e corta (cicli vitali che vanno da 100 a 200 anni) longevità.

Dal punto di vista riproduttivo sono tutte specie monoiche in cui l'autoimpollinazione o la riproduzione agamica, sebbene possano essere importanti in specifici contesti ecologici dove costituiscano un vantaggio evolutivo (Muona & Harju 1989), non sono mai prevalenti sui sistemi di impollinazione incrociata (che generalmente è molto alta con un tasso di esoincrocio superiore al 90%). Nel genere *Pinus* non esistono sistemi di auto-incompatibilità e l'esoincrocio è assicurato principalmente da sistemi di controllo post-embrionale (Kärkkäinen & Savolainen 1993; Koelewijn et al. 1999) o da fenomeni di depressione da inbreeding (o inincrocio) (Sorensen 2001). L'impollinazione è quasi esclusivamente di tipo anemofilo con elevati tassi di flusso genico tra popolazioni (Savolainen & Pyhäjärvi 2007). Non sono infrequenti, infatti, flussi di polline che raggiungono distanze stimate nell'ordine di svariati (centinaia) chilometri (Koski 1970; Schuster & Mitton 2000) promuovendo l'omogenizzazione delle frequenze alleliche tra popolazioni distanti, come è stato dimostrato, ad esempio, tra le popolazioni Svedesi e Cinesi di pino Scozzese (*Pinus sylvestris* L.) che manifestano una ridotta differenziazione tra popolazioni a livello isoenzimatico (Wang et al., 1991). La produzione di seme nei pini raggiunge livelli estremamente abbondanti, con la maggior parte delle specie che presentano sistemi di dispersione mediati dal vento, contribuendo a mantenere una dimensione effettiva della popolazione molto elevata. E' stato dimostrato, in generale, che la dimensione effettiva della popolazione, nella maggioranza delle specie di pini, è molto vicina alla dimensione numerica della stessa (Dodd & Silvetown 2000). I fenomeni dispersivi (flusso genico), oltre a influenzare la struttura genetica spaziale e le dinamiche di espansione degli areali delle popolazioni (Travis et al., 2010), possono contribuire ai processi di adattamento locale (Savolainen et al., 2007; Kremer et al., 2012).

Molte specie, in particolare quelle ad ampia distribuzione, mostrano una variazione clinale di caratteri adattativi (e.g. fenologia, resistenza a stress

abiotici e biotici) mantenuta sostanzialmente dalla selezione naturale (García-Gil et al., 2003; Howe et al. 2003; Chambel et al., 2007). L'esistenza di questi clini genetici in differenti parti dell'areale delle specie è stato osservato in test di provenienza ed esperimenti di campo in condizioni di crescita controllate (i cosiddetti “*common garden approach*”) (Morgenstern 1996), così come in condizioni naturali.

Molteplici indagini tassonomiche sono state realizzate nel genere *Pinus*, basate inizialmente sia sull'analisi dei caratteri morfologici che attraverso studi biochimici (analisi dei terpeni). Questi studi hanno permesso di realizzare una classificazione che prevede la suddivisione delle varie specie in due sottogeneri, *Strobus* (= *Haploxyylon*, *soft pine*) e *Pinus* (= *Diploxyylon*, *hard pine*), in accordo con il numero di fasci cribro-vascolari delle foglie (uno o due); i sottogeneri sono ulteriormente divisi in sezioni e sottosezioni (Gausson 1960; Little & Critchfield 1969; Van der Burgh 1973). L'elevata omoplasia dei caratteri morfologici nei pini ha favorito lo sviluppo di sistemi basati sull'analisi dei polimorfismi a livello del DNA per chiarire le relazioni tra le varie specie. I primi lavori molecolari si sono basati sull'analisi con enzimi di restrizione di regioni del genoma plastidiale e il confronto dei pattern di restrizione tra specie dei due sottogeneri (Strauss & Doerksen 1990; Krupkin 1996), che ha evidenziato la loro origine monofiletica. In seguito, Price e collaboratori (1998) hanno analizzato un elevato numero di specie e una serie di caratteri morfologici, citologici, proteici, alcuni metaboliti secondari e polimorfismi a livello del DNA, proponendo una classificazione che riconosceva 111 specie divise in due sottogeneri, quattro sezioni e 17 sottosezioni. Successivamente sono stati condotti vari studi filogenetici, basati sia su sequenze nucleari ribosomiali (ITS: *internal transcribed spacer*; Liston et al., 1999) che su marcatori plastidiali (*rbcL*, *matK*, *trnV*: Wang et al., 1999; *rbcL* e *matK*: Gernandt et al., 2005), che hanno perfezionato la classificazione riducendo il numero delle sottosezioni, da 17 a 11, e il numero delle specie in accordo con quelle descritte da Farjon (2001).

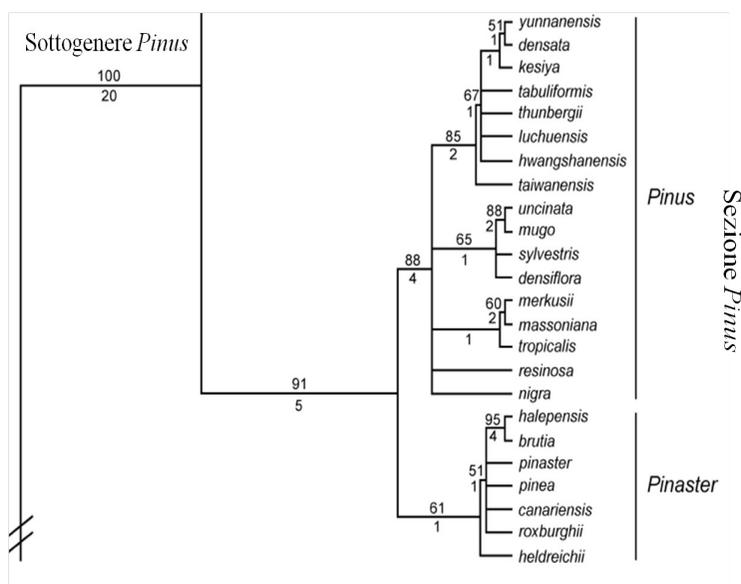
La maggior parte delle specie appartengono al sottogenere *Pinus*, il quale è diviso in due sezioni, la sezione *Pinus*, che comprende principalmente le specie distribuite in Eurasia e in Nord Africa e la sezione *Trifoliae*, che insieme al sottogenere *Strobus*, comprendono quasi esclusivamente le specie distribuite in America (Tabella 1.1).

Tabella 1.1: Sistematica del genere *Pinus*, in accordo con Gernandt et al. (2005).

Sottogenere	Sezione	Sottosezione	numero di specie	Distribuzione
<i>Pinus</i>	<i>Pinus</i>	<i>Pinus</i>	17	Eurasia, Nord America
		<i>Pinaster</i>	7	Mediterraneo, Asia
	<i>Trifoliae</i>	<i>Contortae</i>	4	America
		<i>Australes</i>	26	America
		<i>Ponderosae</i>	17	America
<i>Strobis</i>	<i>Parrya</i>	<i>Balfourianae</i>	3	America
		<i>Cembroides</i>	11	America
		<i>Nelsoniane</i>	1	Centro America
	<i>Quinquefoliae</i>	<i>Gerardiane</i>	3	Asia
		<i>Krempfianae</i>	1	Asia
		<i>Strobis</i>	21	America, Eurasia

La figura 1.1 riporta la struttura filogenetica della sezione *Pinus*, nella quale si possono distinguere due gruppi corrispondenti rispettivamente alla sottosezione *Pinus*, che comprende principalmente i pini distribuiti nell'areale Euroasiatico, e alla sottosezione *Pinaster* che comprende principalmente le specie Mediterranee.

Figura 1.1: Struttura filogenetica del sottogenere *Pinus* utilizzando due marcatori plastidiali (*matK* e *RbcL*) che comprendono 237 siti informativi (Gernandt et al., 2005).



Rimangono da risolvere ancora alcuni problemi di classificazione a basso livello tassonomico e soprattutto all'interno del sottogenere *Strobus* visto che fino ad ora la maggior parte degli studi si sono focalizzati sulle specie appartenenti al sottogenere *Pinus*. Inoltre, le analisi basate sui marcatori plastidiali e sugli spaziatori ITS nucleari risultano poco informative per indagini a basso livello tassonomico, a causa dell'elevato grado di conservazione del genoma plastidiale e dalla rapida natura evolutiva dei marcatori nucleari ITS. La valutazione delle relazioni filogenetiche utilizzando gli spaziatori ITS nucleari, infatti, può produrre risultati non attendibili a causa dell'elevata omoplasia di questi marcatori. A tal proposito indagini molecolari condotte recentemente hanno evidenziato l'elevato contenuto informativo dei loci nucleari ortologi per chiarire le relazioni filogenetiche a basso livello tassonomico (Fulton et al., 2002; Rokas et al., 2003). Il sequenziamento di queste regioni può garantire una migliore risoluzione rispetto alle classiche sequenze plastidiali e nucleari utilizzate in studi di filogenesi e di tassonomia molecolare, per risolvere le relazioni tra specie che si sono differenziate recentemente e per individuare i possibili eventi di ibridazione (Small et al., 2004).

Nei pini, nonostante non esista ancora un genoma di riferimento sequenziato e le dimensioni stimate esibiscono un contenuto di DNA per cellule diploidi tra i più alti del regno vegetale, nell'ordine di 44-75 pg (1 pg = 960 Mb, Arumuganathan & Earle 1991) ovvero 170 volte più grande del genoma di *Arabidopsis thaliana* (Grotkopp et al., 2004), sono stati identificati e mappati numerosi geni e marcatori ESTP (*Express Sequence Tags-Polymorphism*) che hanno mostrato un'elevata ortologia tra le specie analizzate (Brown et al., 2001; Temesgen et al., 2001). Inizialmente questi lavori sono stati utilizzati per identificare loci per creare mappe genetiche di riferimento per il genere *Pinus*. Successivamente, è stato dimostrato che questi marcatori ESTP mostrano elevati livelli di sintenia e collinearità tra le specie prese in esame, il che ha avuto un forte impatto sulla famiglia delle Pinaceae. Ad esempio sono stati intrapresi numerosi studi di genomica comparata che hanno permesso, una volta associato il marcatore ESTP all'ipotetico QTLs, di identificare geni candidati che controllano importanti caratteri. Il mappaggio comparativo dei QTLs tra specie è uno dei principali mezzi per la validazione dei QTL stessi, e rappresenta anche un'importante fase per la loro incorporazione in strategie di miglioramento.

L'utilizzo di approcci multi-locus indipendenti per esplorare le relazioni filogenetiche è solo di recente applicazione nelle conifere (Kusumi et al., 2002); infatti, l'utilizzo di geni nucleari ortologhi è stato ristretto soprattutto alle angiosperme (Cronn et al., 2002) e solo di rado ha visto l'utilizzo nel gruppo dei pini (Syring et al., 2005; Palmè et al., 2009).

Altri lavori hanno utilizzato le informazioni filogenetiche per fare inferenze circa la storia bio-geografica (Eckert & Hall 2006) ed evolutiva (Willyard et al., 2007) del genere *Pinus*, incorporando nei modelli informazioni sui reperti fossili che hanno permesso di determinare i tempi di divergenza tra i taxa analizzati. I risultati ottenuti affermano che la maggior parte delle specie di pini esistenti si è originata recentemente, in particolare quelle che appartengono al sottogenere *Pinus* (14-67 milioni di anni fa) (Willyard et al., 2007). Questi metodi hanno permesso anche di confermare le ipotesi sulla storia bio-geografica ed evolutiva formulate da Millar (1993 e 1998), che prevedono: i) un'origine Mesozoica dei pini alle medie latitudini della Laurasia seguite da una migrazione est-ovest; ii) frammentazione e spostamento dei pini durante tutto l'Eocene; iii) ricomparsa e dominanza dei pini alle medie latitudini nell'Oligocene e Miocene. Le ipotesi di Millar (1993 e 1998) identificano l'Eocene con le sue fluttuazioni climatiche quale periodo più importante nella storia evolutiva dei pini.

L'attuale distribuzione delle specie e la loro diversità genetica è il risultato dei processi di contrazione-espansione delle popolazioni dovute dall'alternarsi di vari cicli glaciali e interglaciali che hanno interessato l'emisfero nord a partire da circa 2,5 milioni di anni fa (Taberlet et al., 1998). Il pool genico delle popolazioni attuali è quindi il risultato degli eventi che hanno fatto seguito all'ultima glaciazione del tardo Pleistocene (130000-10000 anni fa) con variazioni nelle temperature medie annue di 10-14°C, raggiunte in tempi brevi e capaci di perdurare per 50-5000 anni (Dansgaard et al., 1993). Da un punto di vista genetico le specie che occupavano queste nicchie ecologiche sono state sottoposte a fenomeni di erosione della diversità, fino ad arrivare a fenomeni più estremi di estinzione nelle specie meno tolleranti, e successivi fenomeni di “colli di bottiglia” ed “effetto del fondatore” nei processi di ricolonizzazione a partire da piccole popolazioni dislocate in aree rifugio meridionali (Hewitt 1999). Queste aree rifugio sono state individuate grazie a studi filogeografici effettuati in differenti taxa che hanno messo in luce l'esistenza di specifiche aree dove le specie sono

sopravvissute anche durante l'ultimo massimo glaciale (20000-18000 anni fa). In Europa aree rifugio principali sono state localizzate nelle Penisole Iberica, Italiana e Balcanica (in particolare Grecia e Turchia), anche se per specie boreali meno adattabili alle zone Mediterranee è stata ipotizzata l'esistenza di zone rifugio secondarie dislocate in territori più settentrionali (McLachlan et al., 2005; Provan & Bennett 2008).

Un'indagine filogeografica utilizzando marcatori isoenzimatici su 143 popolazioni naturali di *P. sylvestris* campionate su tutto l'areale di distribuzione naturale di questa specie Eurasiatica, ha mostrato che il pool genico delle popolazioni nordiche e centrali è relativamente omogeneo a differenza delle popolazioni orientali e meridionali che sono caratterizzate da livelli significativi di differenziazione. Questo studio ha permesso inoltre di identificare tre rifugi Pleistocenici localizzati nei Balcani, negli Urali meridionali e in Mongolia settentrionale, da cui la ricolonizzazione ha preso origine (Sannikov & Petrova 2012).

Studi filogeografici che coinvolgono diverse popolazioni di pino d'Aleppo (*P. halepensis*), una specie Mediterranea, hanno evidenziato invece, che la diversità genetica neutrale è strutturata lungo un gradiente longitudinale con le popolazioni orientali che ospitano una maggiore diversità e genealogie più antiche rispetto alle popolazioni occidentali (Bucci et al., 1998; Gomez et al., 2001). Questi risultati genetici combinati con dati palinologici e dendroarcheologici suggeriscono uno scenario in cui le popolazioni di pino d'Aleppo si sono espanse verso occidente in tutto il bacino del Mediterraneo da alcuni rifugi localizzati nella parte meridionale della penisola Balcanica, e più precisamente da popolazioni Greche (Climent et al., 2008). Simulazioni demografiche utilizzando sia marcatori plastidiali che geni candidati, oltre a suggerire un'espansione recente (ultimi 10000 anni) della specie, hanno messo in evidenza fenomeni di colli di bottiglia nelle popolazioni occidentali avvenuti subito dopo l'ultima glaciazione (Grivet et al., 2009). Tali simulazioni hanno inoltre confermato i risultati ottenuti utilizzando altri tipi di marcatori molecolari, con popolazioni antiche e stabili in Grecia e popolazioni di recente espansione nelle zone occidentali del Mediterraneo.

In linea generale, per i pini distribuiti nell'areale Mediterraneo sono stati descritti dei clini longitudinali in cui la diversità genetica decresce spostandosi verso occidente e le popolazioni distribuite in questa parte

dell'areale sono geneticamente più omogenee rispetto a quelle distribuite nella parte orientale (Fady 2005).

Gli studi filogeografici hanno evidenziato che, a differenza dei taxa che vegetano in ambienti temperati in cui gli areali di distribuzione delle specie e quindi anche la dimensione effettiva delle popolazioni hanno subito contrazioni rispetto a quelli attuali, le specie Mediterranee non sono state influenzate in maniera significativa dagli eventi di ricolonizzazione e sono state in grado di mantenere dimensioni effettive della popolazione elevata durante le fasi glaciali del Quaternario, in particolare nelle zone orientali del Mediterraneo (Fady & Conord 2010).

Alcuni fattori possono contribuire a rendere meno comprensibili gli scenari filogeografici. Il mescolamento tra diversi pool genici originato da convergenza di rotte di migrazione da rifugi separati può formare zone di contatto, ovvero aree caratterizzate da elevata diversità genetica, che possono essere confuse con le zone rifugio stesse (Petit et al., 2003).

I processi di ricolonizzazione sono influenzati anche dalle barriere biogeografiche. Ad esempio, in Europa la catena Alpina ha agito come un'iniziale barriera all'espansione delle popolazioni provenienti dai rifugi Italiani durante i periodi interglaciali, ritardandone la loro espansione. Le barriere bio-geografiche possano aver contribuito ad incrementare il grado di divergenza tra popolazioni dislocate in differenti zone rifugio, a causa dell'isolamento per distanza. E' il caso, ad esempio, di vari taxa iberici e italiani dove le barriere bio-geografiche presenti (Pirenei, Alpi e mare Mediterraneo) hanno fortemente limitato il flusso genico e le migrazioni da queste aree (Hewitt 1999).

Le osservazioni paleo-palinologiche, sulle quali sono state spesso basate le stime delle velocità di dispersione, hanno in molti casi sovrastimato i tassi di migrazione con valori anche superiori a 1000 m/anno (MacDonald 1993). Stime più recenti, che si originano da modelli che considerano la storia demografica delle specie in esame, hanno proposto per la maggior parte delle specie forestali tassi di migrazione inferiori a 100 m/anno (McLachlan et al., 2005). Appare quindi verosimile supporre che le specie con un areale ristretto siano ancora circoscritte alle zone occupate durante l'ultimo massimo glaciale mentre quelle specie con ampia distribuzione debbano ancora raggiungere una copertura totale degli areali potenziali, nonostante gli ultimi 7000-8000 anni di stabilità climatica.

Molto probabilmente, l'identificazione dei rifugi glaciali e la caratterizzazione delle dinamiche di colonizzazione avvenute nel post-glaciale rappresentano il contributo più innovativo che le indagini filogeografiche abbiano dato agli studi di genetica di popolazione.

Inoltre, la comprensione delle dinamiche passate sulla distribuzione della diversità genetica può essere utilizzata per fare previsioni future sugli impatti demografici delle popolazioni in relazione ai cambiamenti climatici attuali (Pitelka et al., 1997).

La grande quantità di dati di sequenze nucleotidiche che sono state raccolte recentemente grazie all'uso di sistemi di sequenziamento di nuova generazione ha permesso di indagare meglio il ruolo specifico della selezione e della demografia nel definire la distribuzione della diversità genetica in specie forestali. Esistono importanti risorse genomiche per il genere *Pinus*, come le collezioni di EST, mappe genetiche e SNP (vedi Neale & Kremer 2011). Individuare regioni genomiche sotto selezione è uno dei punti fondamentali richiesti per definire le relazioni/associazioni tra genotipo e fenotipo relativo ai caratteri adattativi ed è anche cruciale per comprendere i processi legati all'adattamento (Siol et al., 2010). Uno dei problemi principali quando si studiano questi fenomeni è la difficoltà di distinguere il ruolo relativo della demografia e della selezione. Gli approcci di genomica di popolazione (Luikart et al., 2003), ovvero quelli basati sulla valutazione di un elevato numero di marcatori, sono di recente applicazione su specie non modello, come appunto le conifere, e sono generalmente utilizzati per comprendere le forze evolutive all'origine dell'adattamento, e come queste forze determinano il pattern spazio-temporale della variabilità all'interno e tra le popolazioni.

Un approccio alternativo per studiare gli effetti di questi processi in specie con una storia demografica complessa e con genomi non ancora sequenziati potrebbe essere quello di combinare gli studi basati sui marcatori neutrali (ad esempio marcatori microsatelliti nucleari e plastidiali), che in linea teorica sono influenzati solo dalla storia demografica e non dalla selezione, con gli studi basati sul risequenziamento di geni candidati che codificano per funzioni biologiche potenzialmente correlate con caratteri legati all'adattamento, come quelli relativi alla fenologia (Heuertz et al., 2006; Pyhäjärvi et al., 2007), alla resistenza alla siccità (González-Martínez et al., 2006; Pyhäjärvi et al., 2007; Eveno et al., 2008; Grivet et al., 2009) e alla

resistenza al gelo (Wachowiak et al., 2009). Tutti i loci genomici, infatti, sono influenzati in maniera simile dai processi demografici mentre solo una parte di questi sono interessati dai fenomeni selettivi. Questi studi, che rappresentano anche i primi lavori di ri-sequenziamento su geni candidati in alcune specie di pini, hanno mostrato risultati interessanti. Ad esempio, alcuni autori suggeriscono che colli di bottiglia che si sono verificati prima o dopo l'ultimo periodo glaciale seguiti da un'espansione della popolazione potrebbero in parte spiegare il pattern corrente della variazione molecolare (Pyhäjärvi et al., 2007; Grivet et al., 2009), mentre altri autori hanno chiaramente identificato segni di selezione positiva (Palme et al., 2008; Wachowiak et al., 2009) o selezione bilanciante (Eveno et al., 2008) su alcune specie di pini prese in esame in questi studi. Identificare segni di selezione positiva, che agisce a livello genomico incrementando la frequenza di una mutazione, è di particolare interesse perchè è alla base dell'adattamento locale. Eventi di selezione positiva recenti o ancora in corso possono essere messi in luce utilizzando metodi che si basano sulla valutazione dei polimorfismi all'interno delle specie (Biswas & Akey 2006). Analizzare il pattern della variazione genetica adattativa e trovare associazioni con la variazione fenotipica, stimata attraverso esperimenti di campo, è quindi una nuova sfida. Predizioni teoriche (Le Corre & Kremer 2003) e recenti risultati sperimentali (Eckert et al., 2010) suggeriscono che la variazione genetica adattativa ha una distribuzione pangénomica e che gli approcci multi-locus devono essere implementati per comprendere le basi genetiche della variazione adattativa. Questi approcci prevedono la stima di parametri genetici della popolazione, e.g. diversità nucleotidica e divergenza, e l'esecuzione di test di neutralità accoppiate con inferenze statistiche Bayesiane, su popolazioni campionate su tutto l'areale naturale della specie in esame e utilizzando un elevato numero di marcatori SNP precedentemente identificati in geni coinvolti nell'adattamento.

L'identificazione di geni candidati può essere un processo complicato nel caso del sequenziamento di ESTs con la classica metodologia Sanger, per la difficoltà di sequenziare quelli poco espressi. Questo problema è stato recentemente risolto con l'utilizzo di nuovi approcci, possibili anche nel caso di specie con genomi complessi come quello delle conifere, che sfruttano le tecnologie di sequenziamento di nuova generazione (NGS, *next generation sequencing*), come ad esempio l'RNA-seq, che permettono di ottenere

sequenze anche di trascritti poco abbondanti. Un altro problema è legato alla complessità dei genomi dei pini, per cui i geni per la maggior parte dei caratteri, soprattutto quelli importanti nelle funzioni legate all'adattamento o di interesse commerciale, fanno parte di complesse famiglie geniche (Kinlaw & Neale 1997). Inoltre non bisogna dimenticare che le sequenze promotrici e regolatrici, a monte o a valle del gene (o famiglia genica) considerato, giocano ruoli chiave nella modulazione dell'espressione genica, mentre gli approcci basati sulla valutazione degli EST si concentrano solo sulla porzione codificante e sugli UTR.

Una comprensione approfondita dei fenomeni legati all'adattamento si potranno comunque ottenere confrontando interi genomi. Grazie al rapido sviluppo delle tecnologie NGS, la differenziazione potrà essere monitorata a livello di ampie porzioni genomiche attraverso l'analisi delle divergenze nucleotidiche tra popolazioni contrastanti. Inoltre molte specie di pini formano dei complessi di specie, interfertili e con bassa differenziazione, per cui approcci di scansione genomica potranno permettere l'identificazione e la localizzazione di regioni genomiche responsabili del mantenimento della differenziazione fenotipica delle specie (Neale & Kremer 2011).

Gli approcci basati sul risequenziamento di geni candidati hanno fornito stime di diversità genetica diverse rispetto a quelle ottenute con marcatori neutrali e loci isoenzimatici che mostravano livelli di variazione genetica piuttosto elevati nel genere *Pinus*. Le prime stime della diversità nucleotidica valutate su alcune specie di pini (ad esempio, *Pinus sylvestris*, Garcia-Gil et al., 2003; *Pinus taeda*, Brown et al., 2004; González-Martínez et al., 2006; *Pinus pinaster*, Pot et al., 2005; *Pinus cembra* e *Pinus mugo*, Mosca et al., 2012) evidenziano, a differenza di quanto mostrato precedentemente con l'uso di marcatori neutrali, che le specie appartenenti alle latifoglie (in particolare in alcune specie della famiglia delle querce) mostrano livelli di diversità nucleotidica significativamente più elevati (Savolainen & Pyhäjärvi 2007). Brown e collaboratori (2004) affermano che i bassi livelli di diversità nucleotidica ritrovati in *P. taeda* potrebbe essere dovuto a un tasso di mutazione particolarmente basso (nell'ordine di $0,17 \times 10^{-9}$ /bp/anno, ovvero un ordine di grandezza più basso di quello ritrovato nelle angiosperme, circa $1,5 \times 10^{-8}$ /bp/anno in *Arabidopsis*; Koch et al. 2000).

Conoscere il potenziale genetico adattativo risulta fondamentale in un contesto di cambiamenti climatici, come quelli che si stanno manifestando

ora. Diversi modelli predittivi dimostrano infatti, come il riscaldamento globale indotto dall'uomo sia già una minaccia soprattutto nelle regioni Mediterranee (Giorgi & Lionello 2008). I modelli climatici mostrano come, nell'arco dei prossimi 50-100 anni, le condizioni attuali del clima in Europa subiranno sostanziali cambiamenti che avranno ripercussioni sulla distribuzione delle specie con spostamento verso nord degli areali anche di 100 fino a 400 km. Questo significa che le foreste del sud Europa troveranno condizioni ottimali più a nord, mentre alle latitudini meridionali le specie più sensibili al riscaldamento scompariranno (IPCC 2007a, b). Va comunque sottolineato che questi modelli non assumono la possibilità che si verifichino fenomeni di rapido adattamento alle nuove condizioni climatiche.

Le popolazioni forestali possono adottare tre possibili strategie per far fronte ai cambiamenti climatici: migrare in ambienti più favorevole per la specie; adattarsi *in situ* sfruttando la variazione genetica esistente; estinguersi (Aitken et al., 2008).

Nelle aree montane, in particolare ad altitudini elevate, sia la distribuzione delle specie che la composizione delle foreste sono influenzate dalle condizioni climatiche (Bonan 2008). Un incremento delle temperature potrebbe comportare uno spostamento della distribuzione delle specie ad altitudini più elevate, come è stato già osservato per alcune specie di pini (*P. peuce* Griseb., Meshinev et al., 2000 e *P. mugo*, Camarero et al., 2005). Per la maggior parte delle specie di pini Mediterranei, dove l'eterogeneità topologica di queste regioni e la frammentazione di molte popolazioni costituiscono un vincolo alla migrazione, l'unica possibilità di sopravvivenza a questi repentini mutamenti climatici è l'adattamento *in situ* (Petit et al., 2005).

Per adattarsi *in situ* le specie forestali devono fare affidamento alla variazione genetica già esistente piuttosto che al sorgere di nuove mutazioni, non solo perchè alleli favorevoli sono immediatamente disponibili, ma anche perchè usualmente si ritrovano ad alte frequenze (Innan & Kim 2004). Inoltre un allele già presente (e quindi più "antico") nella popolazione potrebbe risultare più favorevole rispetto ad una mutazione recente in quanto potenzialmente già selezionato ad ambienti del passato, o in un'altra parte dell'areale della specie o anche in un'altra specie con la quale la popolazione ha scambiato geni (Wachowiak et al., 2011).

Individuare la variazione adattativa a livello molecolare, utilizzando marcatori funzionali, è ancora ai primi stadi in specie non modello. Indagini recenti (Grivet et al., 2011) mostrano comunque l'enorme potenziale di questi strumenti nell'identificare popolazioni che ospitano genotipi ben adattati e che potrebbero risultare essenziali per attuare rimboschimenti e più in generale per attuare programmi di conservazione delle risorse genetiche. In futuro ci si attende che le tecniche molecolari basate sui polimorfismi a livello di geni candidati saranno sempre più utilizzate per monitorare i cambiamenti nella composizione genetica delle popolazioni nello spazio e nel tempo (accoppiandole a sistemi basati sul GIS), e per la valutazione dei sistemi di risposta genetica delle foreste ai repentini cambiamenti climatici e ambientali.

1.2 *Pinus nigra*

Il pino nero Europeo (*Pinus nigra* Arnold) rappresenta una delle specie più importanti in Europa meridionale, sia da un punto di vista economico, grazie alla rapidità di crescita nei primi anni di vita e alla qualità del legno elevata, che per progetti di rimboschimento su un ampio spettro di condizioni ecologiche, dovuto all'elevata plasticità adattativa.

Il *Pinus nigra* (Figura 1.2), come le altre specie di pini distribuiti nell'areale Mediterraneo, rappresenta un'entità evolutiva interessante, collegando differenti regioni geografiche nonché differenti discendenze evolutive. Costituisce un'entità eterogenea, in quanto gli eventi paleogeografici del Terziario (in particolare spostamento delle masse continentali e orogenesi delle principali catene montuose Europee) hanno contribuito a frammentare l'areale di distribuzione della specie con un incremento della differenziazione intraspecifica, e consiste principalmente di popolazioni relitte del periodo Cretaceo-Terziario (Klaus 1989; Bogunic et al., 2007).

Figura 1.2: Individui di pino nero (ssp *laricio*) localizzati in una zona montana della Corsica.



E' una specie pioniera eliofila con popolazioni localizzate nella fascia montana, l'ottimo altitudinale va da 800 a 1500 m. La maggior parte delle popolazioni sono adattate a climi di tipo Mediterraneo, ad eccezione delle popolazioni localizzate nella fascia alpina che sopportano meglio climi temperati. Hanno un'elevata capacità di risposta a condizioni climatiche estreme, come la resistenza ai danni da ghiaccio; a tal riguardo le popolazioni distribuite nell'areale orientale della specie resistono a temperature più basse rispetto a quelle distribuite nell'areale occidentale (anche inferiori a -20 °C), ai lunghi periodi di siccità e ai danni da vento.

Il pino nero può crescere in stazioni pure o trovarsi associato ad altre specie di pini come *P. sylvestris*, il *P. mugo*, *P. halepensis* e *P. pinea*. In natura può formare degli ibridi vitali con pino mugo (Schütt 1959), con *P. sylvestris* (Vidakovic 1958) e con *P. leucodermis* (Fukarek 1958).

Da un punto di vista della storia evolutiva il pino nero è una delle specie più antiche del genere *Pinus* in Europa, e discende probabilmente da un gruppo che esisteva già nel periodo geologico del Cretaceo nel sud-est Asiatico (circa 100 milioni di anni fa) (Klaus 1989) che successivamente si è insediato in Europa in seguito a una migrazione verso ovest (Scaltsoyiannes et al., 2009). La presenza della specie in Europa risale al periodo del

Terziario (5-15 milioni di anni fa), in cui ha affrontato la maggior parte dei processi vicarianti di questa regione (come l'ultimo periodo di crisi di salinità del Messiniano, l'orogenesi delle principali catene montuose, e i periodi glaciali-interglaciali del Pleistocene e dell'Olocene) (Mai 1989). La presenza della specie è confermata anche da ritrovamenti fossili datati dal primo Miocene al tardo Pleistocene in Sicilia, sui Pirenei e nel sud della Spagna (Jaramillo-Correa et al., 2010). L'attuale distribuzione (Figura 1.3) della specie non è stata influenzata dall'ultima glaciazione Pleistocenica; uno studio su alcune popolazioni dell'areale occidentale utilizzando marcatori microsattelliti plastidiali conferma con una certa evidenza che queste popolazioni esistevano già durante l'ultimo massimo glaciale nel sud della Spagna, nel sud della Francia, in Corsica e nel sud Italia (Afzal-Rafii & Dodd 2007). Questi autori inoltre, con il supporto di simulazioni demografiche, affermano che non ci sono stati eventi significativi di contrazione/espansione demografica durante l'ultimo periodo glaciale.

Figura 1.3: Areale di distribuzione del pino nero Europeo. (Fonte: EUFORGEN 2009, www.euforgen.org).

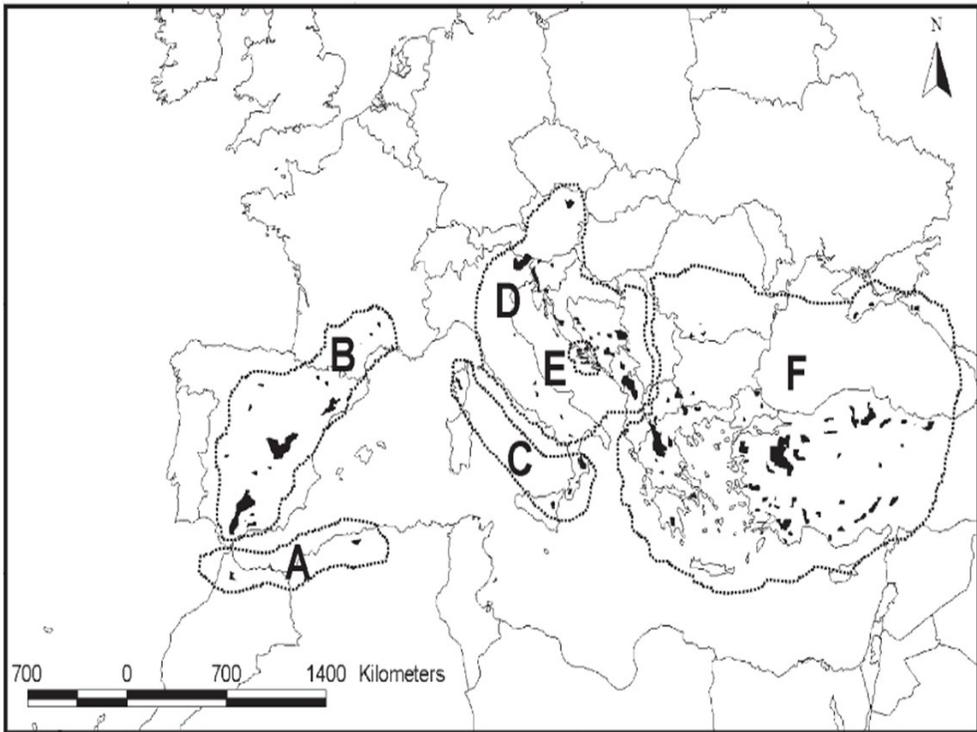


E' una specie ampiamente distribuita nella zona Mediterranea; nonostante questo, il suo areale è estremamente frammentato, principalmente a causa degli eventi avvenuti nel Terziario. Questa distribuzione eterogenea ha causato una divergenza fenotipica elevata anche in popolazioni della stessa regione che ha contribuito enormemente a formulare classificazioni

tassonomiche in disaccordo tra i vari autori, portando a una suddivisione della specie in varie sottospecie e varietà. Gli studi tassonomici sull'intero areale della specie hanno riguardato solo l'utilizzo di marcatori biochimici, come l'analisi della variazione dei terpeni (Gerber et al., 1995; Rafii et al., 1996; Bojovic), mentre gli studi basati sui marcatori molecolari hanno preso in esame un numero limitato di popolazioni e solo raramente la valutazione della diversità genetica ha avuto significato di indagine tassonomica (marcatori isoenzimatici: Fineschi 1984; Aguinagalde & Llorente 1997; marcatori RAPD: Liber et al., 2003; Lucic et al., 2010). Queste informazioni hanno permesso di individuare una struttura bio-geografica in cui possono essere distinte sei sottospecie (Figura 1.4):

- a) *P. nigra* ssp *mauretanica* (Maire et Peyerimh.) presente in Nord Africa, principalmente sulle montagne del Marocco e dell'Algeria;
- b) *P. nigra* ssp *salzmannii* (Dunal) copre un'area estesa della Spagna (dall'Andalusia alla Catalogna), e si ritrova anche sul versante meridionale dei Pirenei e in Francia con alcune popolazioni isolate;
- c) *P. nigra* ssp *laricio* (Poiret) presente nel sud Italia (Calabria e Sicilia) e probabilmente in Corsica;
- d) *P. nigra* ssp *nigra* (Host.) si ritrova in tutto l'arco Alpino, nel nord-est Italia, in Slovenia, Austria e probabilmente parte dei Balcani;
- e) *P. nigra* ssp *dalmatica* (Vis.) si ritrova nei Balcani (Croazia, Bosnia, Serbia e Albania);
- f) *P. nigra* ssp *pallasiana* (Lamb.) copre la porzione più estesa dell'areale e si ritrova in Turchia, Grecia, Cipro e Crimea.

Figura 1.4: Struttura bio-geografica delle sei sottospecie. Le lettere maiuscole identificano le sottospecie e la relativa corrispondenza è riportata nel testo (Fonte: Liber et al., 2003).



Questa suddivisione, che attualmente è la più utilizzata, deve essere comunque confermata da ulteriori indagini molecolari, ad esempio genotipizzando loci SSR nucleari e/o SNP in diverse popolazioni che comprendono tutte le sottospecie.

La diversità genetica in questa specie è stata valutata, fino ad ora, utilizzando quasi esclusivamente marcatori isoenzimatici (Nikolić & Tucić 1983; Tolun et al., 2000; Scaltsoyiannes et al., 2009), mentre solo recente è l'utilizzo di marcatori microsatelliti plastidiali (Afzal-Rafii & Dodd 2007; Jaramillo-Correa et al., 2010; Soto et al., 2010) e ISSR (Moraga et al., 2012). Entrambi i marcatori evidenziano una diversità genetica elevata mantenuta soprattutto all'interno delle popolazioni; le sottospecie più divergenti sono risultate *P. nigra* ssp *salzmannii* e *P. nigra* ssp *laricio* mentre quelle distribuite nell'areale orientale appaiono abbastanza simili, anche se il confronto ha coinvolto un numero limitato di popolazioni. La diversità genetica neutrale (cpDNA) in

questa specie è risultata molto simile a quella di *P. sylvestris*, una specie montana Eurasiatica ad ampia distribuzione e molto più elevata rispetto ad altre specie termofile Mediterranee come *P. pinea* e *P. halepensis* (Vendramin et al., 2008; Soto et al., 2010). Non sono presenti in questa specie stime sulla diversità nucleotidica che possano confermare i risultati ottenuti attraverso l'utilizzo di marcatori neutrali. Come per le indagini a livello delle sottospecie anche questi risultati vanno confermati con marcatori molecolari nucleari (nuSSR e SNP): considerata la storia evolutiva di questa specie, dovremmo attenderci un'elevata differenziazione tra le popolazioni e un alto numero di alleli rari.

2. SCOPO DELLA TESI

I principali obiettivi del lavoro di tesi sono stati:

- a) Generare un dataset di qualità elevata di geni candidati ortologhi in *Pinus nigra*. A tal proposito sono stati considerati 265 geni provenienti da un lavoro di sequenziamento di ESTs in *Pinus taeda* e successive selezioni per un ipotetico coinvolgimento in processi legati all'adattamento sulla base sia di omologie trovate in banca dati che di ricerche bibliografiche. E' stata valutata la percentuale di trasferimento tra queste due specie filogeneticamente distanti e la quantità di geni polimorfici.
- b) Risequenziare 12 geni candidati altamente polimorfici in un campione di 84 individui appartenenti a sette popolazioni di *Pinus nigra* distribuite nell'areale Mediterraneo della specie;
 - b1) Identificare e validare gli SNPs nei geni candidati;
 - b2) Stimare, utilizzando i marcatori molecolari SNPs, la distribuzione della diversità nucleotidica entro e tra le popolazioni di *Pinus nigra* e valutare come questa diversità genetica è distribuita lungo il gradiente geografico della specie;
 - b3) Stimare la differenziazione a livello delle popolazioni attraverso l'analisi molecolare della varianza (AMOVA) e metodi di classificazione Bayesiani (STRUCTURE e TESS) e valutare se la diversità genetica è principalmente strutturata lungo un gradiente longitudinale o è mantenuta a livello delle sottospecie;
 - b4) Implementare dei test di neutralità per identificare geni che discostano dal modello neutrale e fare inferenze sul ruolo relativo della demografia (espansione-contrazione delle popolazioni) e della selezione sulla distribuzione di questi polimorfismi;
 - b5) Verificare a livello tassonomico la possibile presenza di sottospecie attraverso metodiche di classificazione Bayesiana che confermino lo status di complex del *Pinus nigra*.

3. MATERIALI e METODI

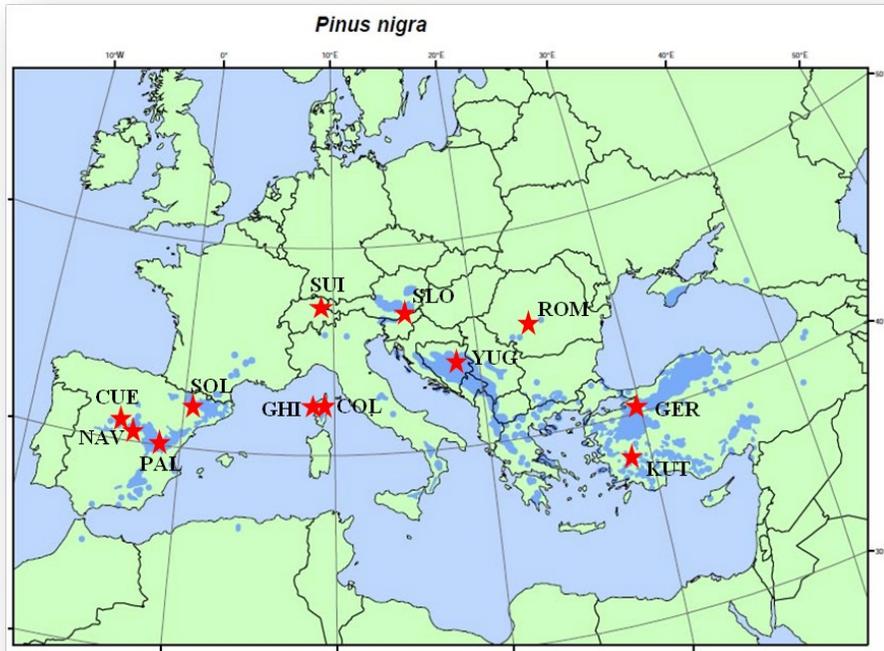
3.1 Generazione di un dataset di geni candidati ortologhi.

In questa prima è stato creato un dataset di geni candidati della specie *Pinus nigra*, in cui si sono controllate e corrette manualmente le sequenze, gene per gene. Il campionamento, l'estrazione del DNA, l'amplificazione dei geni candidati ed il loro sequenziamento è stato realizzato nell'ambito di un progetto finanziato dalla comunità europea ("CRIEC", conifer resequencing initiative).

3.1.1 Campione utilizzato.

In una prima fase, per la creazione del dataset di geni candidati sono stati utilizzati 12 individui (estraendo il DNA da tessuto aploide del megagametofito) di *Pinus nigra* appartenenti a 12 popolazioni distribuite nell'areale naturale della specie (Figura 3.1). Il campionamento è stato realizzato su un ampio gradiente geografico e quindi anche ecologico della specie. I 12 individui comprendono quattro popolazioni localizzate in Spagna, due in Francia (Corsica), una in Svizzera, una in Slovenia, una in Serbia, una in Romania e due in Turchia (Tabella 3.1).

Figura 3.1: Localizzazione delle popolazioni campionate. La sfumatura azzurra sulla mappa identifica l'areale naturale di distribuzione del pino nero Europeo (Fonte: EUFORGEN 2009, www.euforgen.org).



3.1.2 Geni candidati analizzati.

Sono stati considerati 265 geni provenienti da un lavoro di sequenziamento di ESTs in *Pinus taeda* e successive selezioni per un ipotetico coinvolgimento in processi legati all'adattamento (resistenza a stress biotici e abiotici, fenologia della gemma) sulla base sia di omologie trovate in banca dati che di ricerche bibliografiche (Neale et al., dati non pubblicati; Genbank accession da FJ043059 a FJ147084). Il sequenziamento è stato eseguito, in entrambe le direzioni (*forward* e *reverse*), su DNA aploide estratto dal megagametofito, il tessuto di origine materna che circonda l'embrione del seme.

Tabella 3.1: Nome della popolazione, codice identificativo e provenienza degli individui analizzati.

Popolazione	ID popolazione	Provenienza
Solsones	SOL	Spagna
Los Palancares	PAL	Spagna
Navahondona	NAV	Spagna
Cuellar	CUE	Spagna
Ghisoni	GHI	Francia
Col de Vergio	COL	Francia
Suiza	SUI	Svizzera
Slovenia	SLO	Slovenia
Yugoslavia	YUG	Balcani
Rumania	RUM	Romania
Kutahya	KUT	Turchia
Gerede	GER	Turchia

3.1.3 Analisi dei dati.

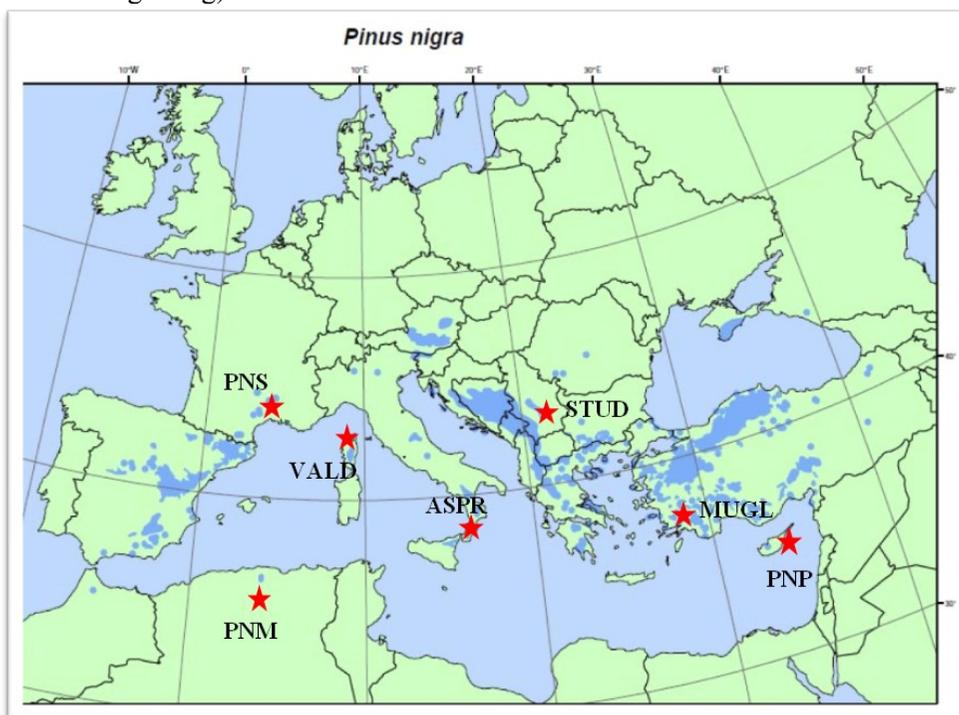
Per ogni gene, le sequenze *forward* e *reverse*, di ogni individuo, sono state assemblate con l'ausilio del *software* Codon Code Aligner (Codon Code Corporation, Dedham, MA, USA) e le sequenze dei 12 individui allineate con MUSCLE (Edgar 2004). Le sequenze di bassa qualità sono state eliminate. L'editing sui multi-allineamenti è stato eseguito anche con il *software* opensource Bioedit 7.0.0 (Ibis Biosciences, Carlsbad, CA, USA) con il quale è stata rilevata la presenza di polimorfismi a singolo nucleotide (SNPs) e sono stati salvati tutti gli allineamenti in formato "*fasta*". Sulla base del dataset contenente solo sequenze di alta qualità, è stata valutata la frequenza di trasferimento in *P.nigra* dei geni candidati, i cui primers sono stati disegnati sulla base delle sequenze di cDNA di *P.taeda*. In aggiunta, le sequenze sono state analizzate per la possibile presenza di regioni microsatellite (ripetizioni di- tri- e tetra-nucleotidi, con motivo ripetuto ≥ 8), utilizzando il *software* on-line Sputnik (<http://www.cbib.u-bordeaux2.fr/pise/sputnik.html>). E' stato inoltre determinata la percentuale di geni che hanno manifestato polimorfismo tra i 12 individui campionati.

3.2 Risequenziamento di 20 geni candidati in sette popolamenti di *Pinus nigra*.

3.2.1 Campione utilizzato.

Il campione utilizzato nella fase di ri-sequenziamento è formato da sette popolazioni distribuite nell'areale di distribuzione Mediterraneo della specie (Figura 3.2). Ogni popolazione è composta da 12 individui diploidi (N tot. = 84).

Figura 3.2: Distribuzione delle popolazioni utilizzate per il risequenziamento dei 20 geni selezionati. La sfumatura azzurra sulla mappa identifica l'areale di distribuzione naturale del pino nero Europeo. (Fonte: EUFORGEN 2009, www.euforgen.org).



La tabella 3.2 riassume alcune informazioni inerenti alle sette popolazioni quali: il nome e il codice identificativo della popolazione, la sottospecie di appartenenza, la loro provenienza e le coordinate geografiche.

Il campionamento del materiale vegetale e l'estrazione del DNA, diploide visto che proviene da aghi, è stata eseguita da ricercatori che fanno parte del consorzio CRIEC.

Tabella 3.2: Informazioni sulle sette popolazioni incluse nell'analisi di risequenziamento.

Popolazione	ID POP	Sottospecie	Provenienza	Lat (N)	Long (E)
Djurdjura	PNM	<i>mauretanica</i>	Algeria	36°27'	4° 06'
Aspromonte	ASPR	<i>laricio</i>	Italia	38°14'	15°59'
Valdoniello	VALD	<i>laricio</i>	Francia	41°51'	9°06'
Saint Guilhem	PNS	<i>salzmanni</i>	Francia	43°46'	3°34'
Studenica	STUD	<i>nigra o dalmatica</i>	Serbia	43°29'	20°32'
Mugla	MUGL	<i>pallasiana</i>	Turchia	37°14'	28°22'
Cyprus	PNM	<i>pallasiana</i>	Cipro	34°57'	32°50'

3.2.2 Scelta e amplificazione dei geni candidati

Sono stati scelti 20 geni candidati tra quelli selezionati nella prima fase (sequenziamento di 265 geni in 12 individui di 12 popolazioni) sulla base dei livelli di polimorfismo e la qualità delle sequenze. Le 20 coppie di primers utilizzate sulle sette popolazioni di *Pinus nigra* sono elencate in Tabella 3.3.

Tabella 3.3: Primer considerati per il risequenziamento di 7 popolazioni di Pino nero Mediterraneo.

Codice geni	Primer forward	Primer reverse
0_10162_01	GTGGTGAGAAAATGGG	AGAGGTAACAAAGCGG
0_10384_02	CATGGGAAGAAGGATTTAC	TGCGGTTGTATGCTTG
0_10667_02	TTATATCAAAAGCGTGCTGG	CTGGATGACATAGAAGTTGG
0_12216_02	AAGCAAATGAATACTGGGGAC	TCTTGCTCAGCACAAATAATG
0_13484_01	ATTCATCCACTTCCCCC	GTGTATTTGGTGTGCAG
0_13957_02	ACATGGTGTTTGGCAGG	AATGAGGAGGGTGGTAGAG
0_14221_01	GATCATTCTAGGCACAGCACAAG	CCTGTACCGTGTTCATCAATTTAGCAAG
0_16068_01	GAAGGCAGTTTTGGAAG	GGGAACAGCCAACTATATG
0_16810_02	CCATCTATTCCCAGTG	TCCTTCTAAAGTCGCC
0_18101_02	TTACCGTTTCACTGTGAACTTG	AGAGAGTTTTCCCTTTGTCTGATG
0_18644_02	TATCTTCACACAAACTGCC	CTTTTCTACTGACCATTCCCC
0_2078_01	TCGAAGACAAGCTCTG	GACGACGATAAATGCTAC
0_4229_01	CTGAGAAGAAAGGCTGGATTG	ATTCGAGGTGGCTGCGGAAATG
0_5575_01	TACAATCCTCCCCAAC	CTGTTTCAGCCCATC
0_6293_01	CTATCATCTTGCCTGGGCTC	CAGCTACCCCTCTTAAAGCAGAAAAAC
0_7916_01	GAAGACAACCATAACTACAG	CAGAGTTTTCAAAAATAGCCAGGAC
0_8479_01	CTGCTCAAGTGAAAGG	CTGATTGTGGATTCTGTG
0_9383_01	TAGAGAGCAGATGGAG	ATGGAGCACAGCAACAG
2_1405_01	CATTTTGAGAGGCAAG	TGCGTAAGGCAGAACAG
CL4470Ct1_01	CCTCATCTACCCATATTAC	GATCCAGACAGACATGCAG

La reazione di amplificazione è stata eseguita su un volume totale di 14 μ L contenente: 20 ng di DNA genomico, 1X di buffer (Promega Corp. Madison, USA), 0.2 mM di dNTP (Promega), 0.05 U di Taq polimerasi (Promega), 0.2 μ M di primer *forward* e 0.2 μ M di primer *reverse*. La PCR è stata realizzata in un termociclatore dell'Applied Biosystem (AB System, Germania). Tutte e 20 le coppie di primers sono state amplificate con il seguente profilo termico: 94°C per 3 min, 10 cicli *touch-down* a 94°C per 30 sec, 60°C per 30

sec con un decremento di 1°C ogni ciclo e 72°C per 40 sec, seguiti da 25 cicli a 94°C per 30 sec, 50°C per 30 sec e 72° per 40 sec, con un'estensione finale di 72°C per 7 min. E' stato eseguito, per ogni gene analizzato, anche un controllo negativo, escludendo il DNA dalla mix di reazione. Per valutare la presenza del prodotto di amplificazione delle dimensioni attese, è stata effettuata un'elettroforesi su un gel di agarosio all'1,5%.

La purificazione dei prodotti di PCR e il loro sequenziamento, in entrambe le direzioni, è stato eseguito presso IGA Technology Services srl (Udine).

Gli stessi individui sono stati genotipizzati ad alcuni loci microsatellite, più precisamente a quattro loci nucleari e cinque plastidiali.

3.2.3 Analisi dei dati.

Per ogni gene, le sequenze *forward* e *reverse*, di ciascun individuo, sono state assemblate con l'ausilio del *software* Codon Code Aligner (Codon Code Corporation, Dedham, MA, USA) e allineate con MUSCLE (Edgar 2004). E' stato quindi effettuato "editing" manuale delle sequenze allineate, eliminando quelle relative ai primers, e cercando e validando gli SNPs. La validazione degli SNPs in sequenze che derivano da materiale diploide è una fase critica perché l'eterozigote è visualizzato negli elettroferogrammi come doppio picco sovrapposto, il che aumenta la probabilità di errori nell'identificazione di SNPs (vedi Figura 3.3). Per ridurre questo possibile errore, è stata utilizzata una soglia di "Phred quality" superiore a 30.

Figura 3.3: Multi-allineamento di un gene candidato in alcuni individui di *P.nigra*. Viene evidenziato uno SNP eterozigote con il rispettivo elettroferogramma che mostra chiaramente la presenza di due picchi sovrapposti corrispondenti all'allele materno e paterno (C/G).



La fase delle sequenze diploidi è stata ricostruita con l'algoritmo PHASE 2.1 (Stephens et al., 2001) che utilizza un metodo Bayesiano basato sulla coalescenza per inferire gli aplotipi e produrre due sequenze aploidi per individuo. Le inserzioni/delezioni ("INDEL") sono state codificate come SNPs.

L'analisi dei dati di sequenza è stata fatta utilizzando il software DnaSP v5.0 (Librado & Rozas 2009). La diversità nucleotidica è stata valutata stimando il θ_w di Watterson (Watterson 1975) e il π , il numero medio di differenze tra coppie nucleotidiche in un campione di sequenze (Nei 1987).

E' stata quindi condotta una serie di analisi statistiche, per ogni sequenza genica, al fine di identificare lo scostamento dal modello sull'evoluzione neutrale, come il D di Tajima (1989), il test di Fu e Li (1993) e il test Fs di Fu (1997). La statistica D di Tajima è stata calcolata per ogni locus e riflette le differenze tra π e θ_w . I test di Fu e Li (D e F) sono stati calcolati senza l'outgroup, e considerano la distribuzione delle mutazioni nella genealogia confrontando le mutazioni "nuove" con quelle "vecchie". Il test neutrale Fs è basato sulla distribuzione delle frequenze aplotipiche. I valori di D e Fs e la loro significatività sono stati calcolati anche con il *software* Arlequin 3.5. (Excoffier & Lischer 2010).

L'analisi della varianza molecolare (AMOVA) è stata stimata con il *software* Arlequin 3.5. (Excoffier & Lischer 2010). Il calcolo dell'AMOVA (Excoffier et al., 1992) è stato eseguito considerando tre diversi livelli gerarchici: i) un solo gruppo che comprende tutte e 7 le popolazioni; ii) due gruppi suddivisi in popolazioni occidentali e orientali; iii) cinque gruppi che comprendono le cinque sottospecie. La significatività degli indici di fissazione è stata stimata attraverso 1000 permutazioni.

Con tutti i geni concatenati, è stato creato un dataset multi-locus formato da 270 loci polimorfici (SNPs e indel) sul quale è stata valutata la possibilità di identificare loci che manifestano segnali di selezione attraverso il *software* LOSITAN (Antao et al., 2008) che implementa un metodo basato sulla ricerca di valori estremi di Fst (*Fst-outlier method*). Questo metodo valuta la relazione tra coefficiente di differenziazione (Fst) e eterozigosi attesa (H_{exp}) in un modello di migrazione casuale tra popolazioni separate (*island model*), simulando una distribuzione attesa dei due indici sotto un modello neutrale. Questa distribuzione è utilizzata per identificare i loci che possiedono valori significativamente più alti o più bassi di Fst rispetto a quanto atteso secondo

una distribuzione neutrale. Per calcolare la distribuzione, l'intervallo di confidenza dello 0,99 e un FDR (*False Discovery Rate*) dello 0,1 sono state eseguite 50000 simulazioni.

La struttura delle popolazioni è stata analizzata eliminando dal dataset quei loci che sono risultati sotto selezione secondo l'analisi degli *Fst-outlier*. La struttura genetica è stata valutata attraverso il *software* STRUCTURE v2.3.4 (Pritchard et al., 2000; Falush et al., 2003) che implementa un algoritmo Bayesiano di raggruppamento non spaziale e considera l'equilibrio di Hardy-Weinberg e il *linkage disequilibrium*. L'informazione geografica può essere incorporata a priori nei modelli analizzati oppure utilizzata a posteriori per il controllo delle relazioni con la struttura genetica stimata. È stato utilizzato modello di *admixture*, nel quale, per ogni individuo è stimata la frazione di ascendenza a ciascun cluster con le frequenze alleliche correlate.

Sono state implementate tre corse indipendenti per ciascun K (cluster o gruppo) con un valore di K che va da 1 a 7, un *burn-in-period* di 50000 steps seguiti da 100000 simulazioni MCMC (Markov Chain Monte Carlo). Per verificare la robustezza dei risultati delle differenti corse per un determinato valore di K, è stato calcolato il coefficiente di similarità (SC) fra coppie di corse, come descritto da Rosenberg e collaboratori (2002). Inoltre per verificare il numero di gruppi (K) che meglio spiega i dati è stato calcolato il deltaK (Evanno et al., 2005). Il calcolo del deltaK e del coefficiente di similarità è stato eseguito con il software R e il pacchetto STRUCTURE-SUM (Erich 2006) (<http://taylor0.biology.ucla.edu/structureHarvester/>). Per valutare la possibile presenza di sottostrutture all'interno di ciascun gruppo, è stata indagata la struttura all'interno dei gruppi utilizzando gli stessi parametri di simulazione descritti sopra ad eccezione del numero di K che varia in funzione del numero di popolazioni che compongono i gruppi. Per ogni raggruppamento ottenuto con STRUCTURE sono state stimate sia la diversità che la differenziazione genetica (*Fst*) tramite il software Genetix 4.04 (Belkhir et al., 2000).

I dati sono stati anche analizzati con un altro *software* di *clustering* Bayesiano che considera anche la distribuzione spaziale delle popolazioni, TESS v2.3 (Chen et al., 2007). Il programma tenta di ricostruire la struttura genetica dai genotipi multi-locus individuali, campionati in distinte località senza assumere nessuna popolazione predefinita, e definisce la probabilità di appartenenza e l'assegnamento degli individui al loro cluster geografico. È

stato utilizzato modello *admixture* (Durand et al., 2009) implementando 5 corse indipendenti per ciascun K, con il valore di K che varia da 1 a 10, un *burn-in-period* di 10000 steps seguiti da 50000 interazioni. Per identificare il miglior K (Kmax), i valori di DIC (*Deviance Information Criterion*) delle repliche per il cluster considerato sono stati mediati e quindi riportati in grafico in funzione del numero di clusters: il miglior K corrisponde a quello per cui si osserva un andamento del DIC che va a plateau. Il DIC è una misura statistica di predizione del modello, in principio modelli con DIC piccoli sono preferiti. Tramite il programma CLUMPP (Jakobsson & Rosenberg 2007) è stata calcolata la probabilità di appartenenza media, tra le repliche, dei vari cluster in modo da poter utilizzare questi valori nelle rappresentazioni spaziali (ad esempio utilizzando il *kriging*).

La struttura genetica è stata valutata anche con l'utilizzo di marcatori neutrali quali microsatelliti nucleari e plastidiali (quattro loci microsatelliti nucleari e cinque loci microsatellite plastidiali). I loci SSR nucleari sono stati analizzati con entrambi i *software* di clustering Bayesiana (STRUCTURE e TESS) mentre per quelli plastidiali è stato utilizzato solo TESS, visto che le assunzioni di STRUCTURE (*HW equilibrium* e *linkage equilibrium*) non si applicano a marcatori ad ereditarietà uniparentale. I parametri utilizzati sono stati gli stessi descritti sopra, sia per STRUCTURE che per TESS, così come il numero di K che varia in funzione del numero di gruppi considerati.

Al fine di indagare la classificazione gerarchica delle popolazioni, sono state calcolate le distanze genetiche di Nei, Da, (Nei 1987) attraverso il *software* Population 1.2.28 (Langella 2002). La matrice delle distanze genetiche è stata usata come input file per il *software* TreeExplorer 2.12 (Tamura, 1999) per costruire il dendrogramma con l'algoritmo *neighbor joining* (NJ). E' stato utilizzato un valore di *bootstrap* di 1000 repliche, calcolato con il *software* Population 1.2.28.

La rappresentazione spaziale delle distanze genetiche tra le popolazioni è stata eseguita con GenGIS ver2.0 (Parks et al., 2009), un'applicazione bioinformatica che permette di combinare le informazioni spaziali con i dati molecolari. La mappa geo-referenziata delle popolazioni è stata ottenuta con l'ausilio di MapMaker ver1.0, un programma compatibile con GenGIS che permette di utilizzare le mappe mondiali ad alta qualità.

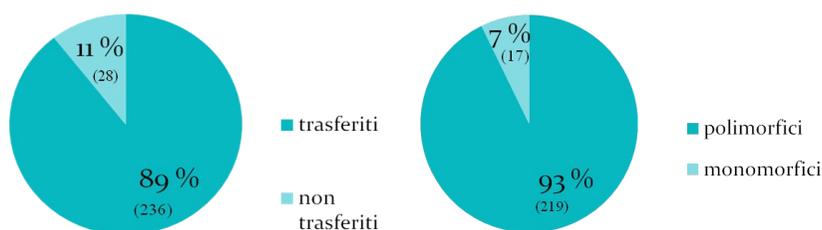
4. RISULTATI

4.1 Generazione di un dataset di geni candidati ortologi

E' stato generato un dataset di geni candidati ortologi di qualità elevata, i cui primers sono stati sviluppati in *Pinus taeda*, una specie di pino nativo del Nord America.

Su 264 geni valutati inizialmente, ne sono stati eliminati 28 a causa della bassa qualità delle sequenze ottenute. Dei 236 geni trasferiti con successo, 219 sono risultati polimorfici e i restanti 17 monomorfici (Figura 4.1).

Figura 4.1: Trasferibilità dei geni candidati trasferiti e loro polimorfismo.

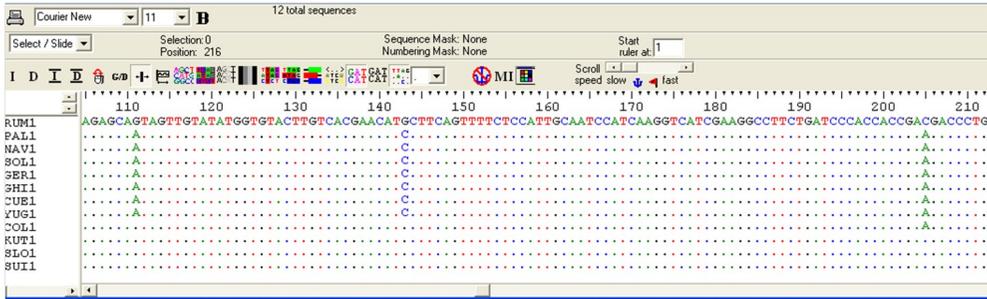


La lunghezza media delle sequenze è di circa 500 bp. Sono stati valutati circa 5000 cromatogrammi per un totale di oltre 1 milione di bp.

La figura 4.2 mostra l'allineamento tra i 12 individui di pino nero, appartenenti ciascuno a una popolazione distinta, di una sequenza genica dove sono presenti tre SNP fissati in alcune popolazioni all'interno della specie.

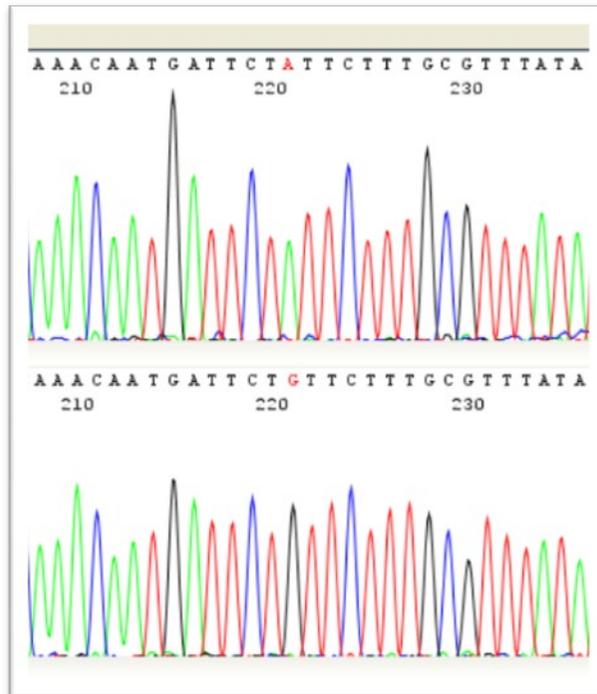
Queste sequenze derivano dal DNA estratto dal megagametofito, il tessuto aploide di origine materna che circonda l'embrione, per cui l'identificazione degli SNP non è ambigua (Figura 4.3).

Figura 4.2: Allineamento di una sequenza genica tra i 12 individui di *Pinus nigra*.



Non è stata identificata nessuna regione microsatellite (di-, tri- e tetra-nucleotide) con le caratteristiche riportate nei materiali e metodi, nelle sequenze geniche analizzate.

Figura 4.3: Elettroferogramma di una sequenza che mostra un polimorfismo a singolo nucleotide (SNP) in due popolazioni di *P.nigra*.

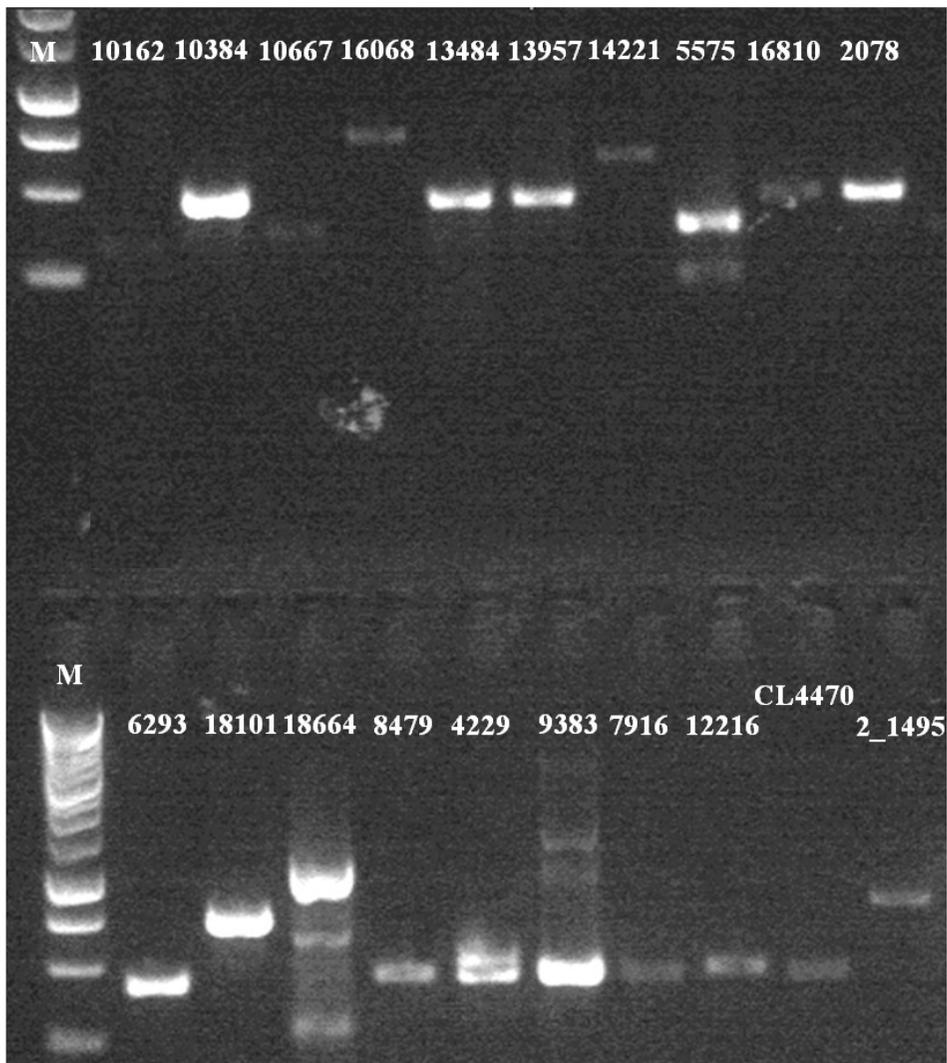


4.2 Risequenziamento di 20 geni candidati in sette popolamenti di Pinus nigra.

In una prima fase 20 coppie di primers, che amplificano 20 geni candidati selezionati dal dataset precedente, sono state valutate per la possibilità di generare un'amplificazione a singola banda su un campione rappresentativo appartenente a tutte e sette le popolazioni campionate nella fase di risequenziamento (Figura 4.4). Successivamente in base ai risultati ottenuti sono stati selezionati 16 geni che sono stati amplificati su tutti gli individui. Il gene 0_16068_01 è stato scartato perché pur presentando una singola banda la dimensione osservata dell'amplificato sul gel di agarosio era molto superiore alla dimensione attesa; analogamente altri tre geni che presentavano un pattern di amplificazione multibanda: 0_18664_02, 0_5575_01 e 0_9383_01. Altri due geni sono stati eliminati durante l'editing delle sequenze: il gene 0_4229_01 presentava delle sequenze di bassa qualità (Figura 4.5) mentre il gene 0_12216_02 manifestava polimorfismi eterozigoti fissati in tutte le popolazioni (Figura 4.6), probabilmente a causa di un gene paralogo.

Tutte le analisi genetiche sono state effettuate su un totale di 14 geni candidati.

Figura 4.4: Controllo su gel di agarosio delle dimensioni attese dei 20 geni testati nelle analisi preliminari.



Note: M rappresenta il marker di peso molecolare noto (1kbase).

Figura 4.5: Sequenza genica di bassa qualità scartata nelle analisi preliminari.

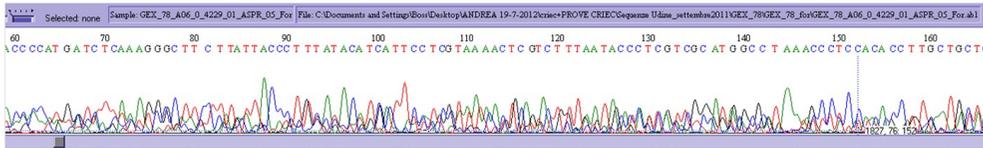


Tabella 4.1: Statistiche descrittive di diversità nucleotidica nei 14 geni selezionati (mediate su tutte le popolazioni di *Pinus nigra*).

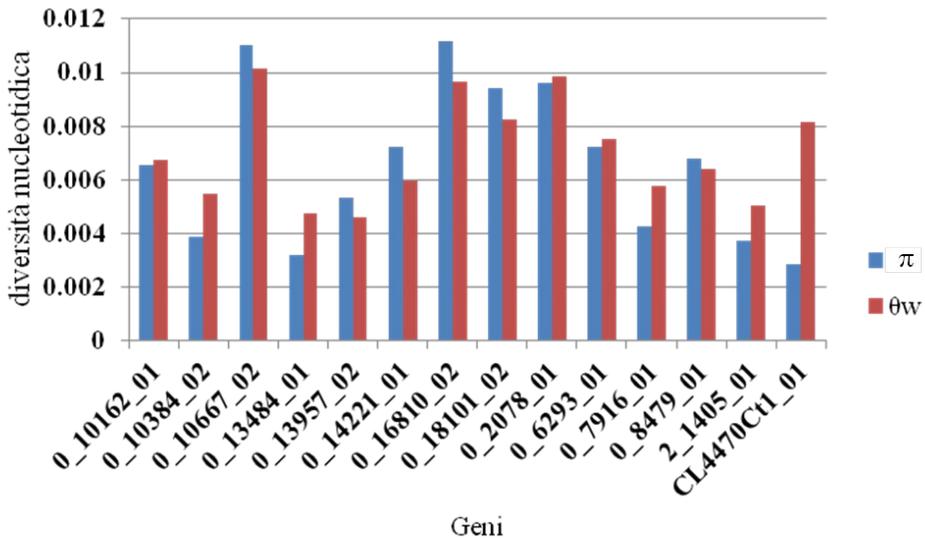
Gene	N	Lunghezza		
		(bp)	indel	S
0_10162_01	78	446	23	16 (2)
0_10384_02	83	417	-	13 (1)
0_10667_02	83	317	5	18 (1)
0_13484_01	36	436	-	10 (1)
0_13957_02	53	415	1	10 (2)
0_14221_01	84	559	-	19 (1)
0_16810_02	77	389	2	21 (5)
0_18101_02	80	666	-	31 (0)
0_2078_01	75	404	5	22 (3)
0_6293_01	83	328	-	14 (0)
0_7916_01	81	398	-	13 (1)
0_8479_01	82	552	1	20 (3)
2_1405_01	71	754	-	21 (2)
CL4470Ct1_01	81	544	1	25 (1)

Gene	h	Hd	k	π	θ_w
0_10162_01	30	0,923	2,765	0,00654	0,00673
0_10384_02	13	0,520	1,610	0,00386	0,00548
0_10667_02	24	0,911	3,441	0,01103	0,01015
0_13484_01	7	0,676	1,387	0,00318	0,00473
0_13957_02	15	0,780	2,203	0,00532	0,00461
0_14221_01	24	0,869	4,034	0,00722	0,00596
0_16810_02	16	0,699	4,319	0,01116	0,00967
0_18101_02	46	0,956	6,285	0,00944	0,00824
0_2078_01	30	0,903	3,831	0,00960	0,00987
0_6293_01	26	0,824	2,373	0,00724	0,00751
0_7916_01	22	0,725	1,703	0,00428	0,00577
0_8479_01	36	0,915	3,741	0,00679	0,00640
2_1405_01	33	0,889	2,811	0,00373	0,00504
CL4470Ct1_01	23	0,739	1,544	0,00284	0,00813

Note: **N:** numero di individui analizzati; **lunghezza:** lunghezza in bp delle sequenze; **S:** numero di siti segreganti (in parentesi è riportato il numero di singletons); **h:** numero di aplotipi; **Hd:** diversità aplotipica; **k:** numero medio di differenze nucleotidiche (Tajima 1983); **π :** diversità nucleotidica (Nei 1997); **θ_w :** diversità nucleotidica (Watterson 1975).

La rappresentazione grafica della variazione della diversità nucleotidica valutata dal confronto tra gli stimatori π e θ_w in *Pinus nigra* è mostrata in figura 4.7.

Figura 4.7: Diversità nucleotidica in *Pinus nigra* valutata in 14 geni con gli stimatori π e θ_w .

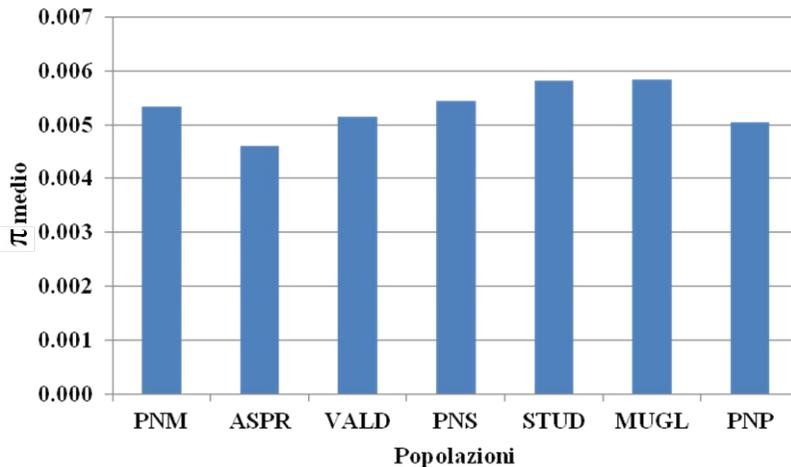


Per la maggior parte delle sequenze geniche le due stime forniscono valori simili di diversità nucleotidica all'interno dello stesso gene, ad eccezione del gene CL4470Ct1_01 dove il valore dello stimatore di Watterson è più del doppio del valore di π . I geni 0_10667_02, 0_16810_02 e 0_2078_01 manifestano i valori più alti.

La variazione della diversità nucleotidica media tra le popolazioni è mostrata in figura 4.8. Le popolazioni appartenenti all'areale orientale della specie (STUD, MUGL e PNP) ospitano una maggiore diversità rispetto alle popolazioni occidentali (PNM, ASPR, VALD e PNS). La bassa diversità genetica della popolazione cipriota (PNP) può essere spiegato dalla ridotta dimensione della popolazione effettiva e dal ridotto flusso genico dovuto all'isolamento geografico di quest'isola. Stesso discorso è possibile fare per le popolazioni Calabrese (ASPR) e Corsa (VALD).

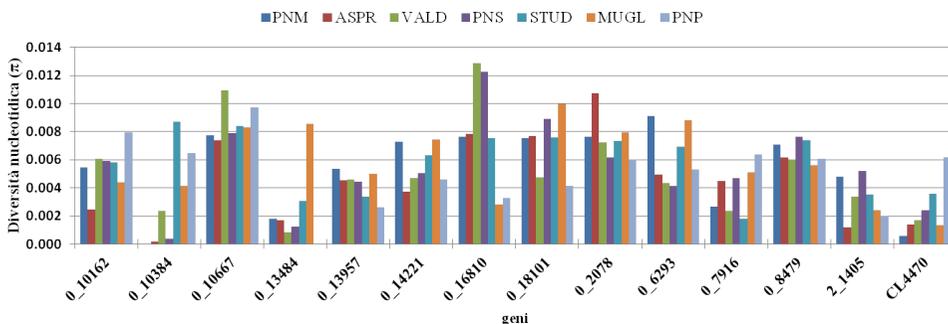
La diversità nucleotidica (π) in *Pinus nigra* calcolata come valore medio sui 14 geni analizzati è risultata pari a 0,0058.

Figura 4.8: Diversità nucleotidica media (π) tra le 7 popolazioni.



La figura 4.9 illustra la variazione nucleotidica tra i 14 geni all'interno delle popolazioni.

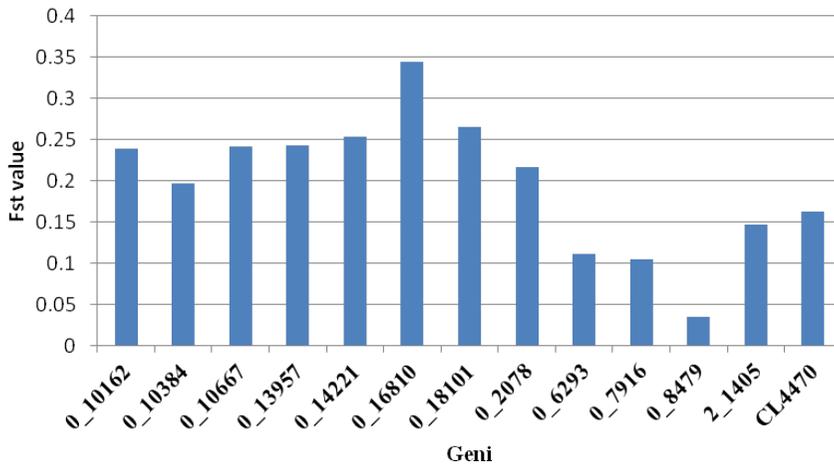
Figura 4.9: Confronto della diversità nucleotidica (π) tra popolazioni e tra i 14 geni.



L'analisi della varianza molecolare è stata eseguita per ogni gene. Il gene 0_13484_01 è stato escluso da questa analisi a causa del basso numero di individui presenti in alcune popolazioni. La figura 4.10 mostra i valori di differenziazione genetica tra popolazioni (F_{st}) tra i vari geni. Tutti i geni manifestano valori di differenziazione genetica significativi; in media il 20%

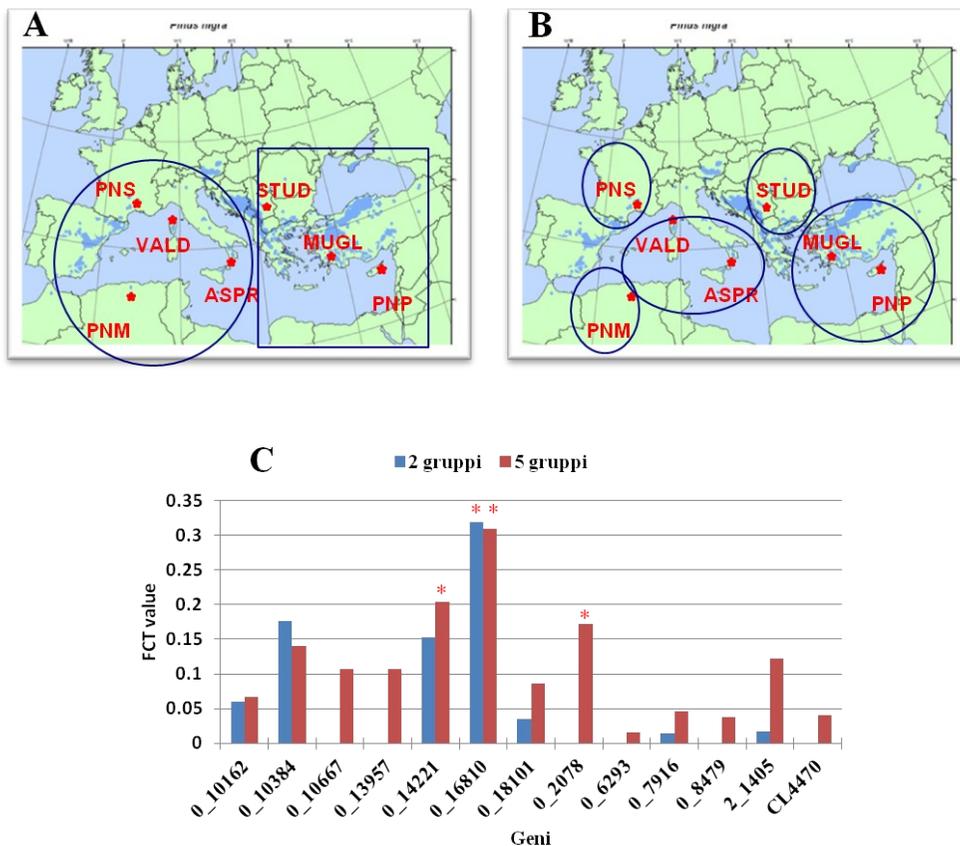
della variazione molecolare è dovuta a differenze tra popolazioni, con valori estremi che vanno da 35% (gene 0_16810_02) al 4% (gene 0_8479_01).

Figura 4.10: Distribuzione dei valori di Fst (differenziazione genetica tra popolazioni) tra i vari geni. Tutti i valori sono altamente significativi ($P < 1\%$).



L'AMOVA è stata eseguita considerando inoltre due modelli di suddivisione delle popolazioni: i) suddividendo le popolazioni in due gruppi che corrispondono alla distribuzione est-ovest dell'areale della specie; ii) suddividendo le popolazioni in cinque gruppi sulla base della loro appartenenza a sottospecie (Figure 4.11A e 4.11B). La figura 4.11C mostra i risultati ottenuti dall'analisi della varianza molecolare confrontando i due modelli. Il modello a due gruppi mantiene in media il 6% della variazione molecolare mentre per il modello a cinque gruppi oltre l'11% di variazione molecolare è dovuto a differenze tra le sottospecie.

Figura 4.11: Analisi della varianza molecolare considerando una suddivisione delle popolazioni in due gruppi (A) e in cinque gruppi (B). I risultati delle analisi sono evidenziate in (C). La significatività degli F_{CT} è indicata da un * ($P < 1\%$).



4.2.2 Test di neutralità.

La tabella 4.2 riassume i risultati dei test di neutralità ottenuti per ogni gene valutando tutto il campione. La maggior parte dei geni non discostano in maniera significativa dall'equilibrio neutrale. Il trend generale del test di Tajima mostra che 8 geni su 14 hanno valori negativi, indicando un eccesso di alleli rari, ma solo il gene CL4470Ct1_01 ha un valore significativamente differente da zero, il che può indicare una possibile selezione positiva che agisce su questo gene. Il trend delle statistiche di Fu & Li (D^* e F^* test) mostra che tutti i geni hanno valori positivi, ad eccezione di 0_16810_02, il che può indicare un difetto di mutazioni recenti. I geni 0_18101_02 e 0_6293_01 hanno valori altamente significativi del test D^* (Fu & Li)

indicando che la mutazione è avvenuta nella parte antica della genealogia, così che gli eventi selettivi o demografici responsabili del pattern osservato sono relativamente antichi. Il trend osservato dei valori del test statistico di Fu (Fs-test) mette in risalto, per quasi tutti i geni, un eccesso di aplotipi rispetto a quanto atteso in accordo con l'ipotesi nulla di neutralità. Otto geni su 14 presentano valori negativi altamente significativi, che può indicare una possibile espansione demografica.

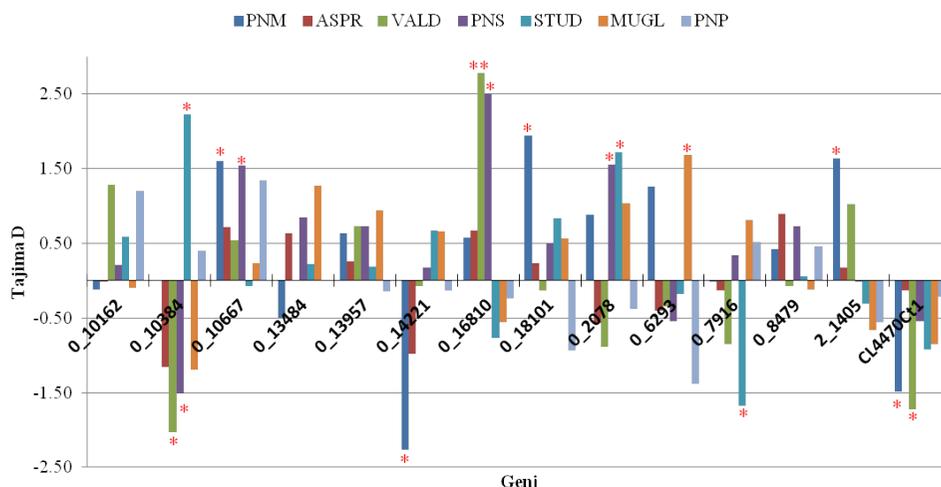
Tabella 4.2: Risultati dei test di neutralità per ogni gene. La significatività del test è indicata da * ($P < 5\%$) e da ** ($P < 1\%$).

Gene	Tajima D	Fu&Li D*	Fu&Li F*	Fu Fs
0_10162_01	-0,495	0,707	0,291	-16,257**
0_10384_02	-0,748	0,843	0,302	-3,202
0_10667_02	0,234	1,155	0,959	-5,99
0_13484_01	-0,880	0,724	0,210	-0,443
0_13957_02	0,389	-0,051	0,123	-0,443
0_14221_01	0,567	1,205	1,149	-4,346
0_16810_02	0,426	-0,581	-0,225	0,216
0_18101_02	0,422	2,011**	1,635	-18,940**
0_2078_01	-0,077	0,439	0,284	-10,922*
0_6293_01	-0,092	1,534*	1,108	-13,067*
0_7916_01	-0,655	0,846	0,347	-12,726*
0_8479_01	0,167	0,269	0,275	-17,942**
2_1405_01	-0,726	0,847	0,283	-21,193**
CL4470Ct1_01	-1,837*	1,459	0,169	-15,654*

I test di neutralità sono stati realizzati anche a livello di popolazioni. La figura 4.12 mostra i valori del test di Tajima, le figure 4.13 e 4.14 riportano i risultati trovati con i test D* e F* di Fu e Li, rispettivamente.

I valori del test di Tajima calcolati tra le popolazioni, mostrano molti geni che discostano in maniera significativa dall'equilibrio neutrale. A differenza dei valori ritrovati quando si considerando tutti gli individui come appartenenti ad un'unica popolazione, si può notare come molti geni abbiano pattern contrastanti tra le popolazioni. Ad esempio il gene 0_10384_02 mostra un valore positivo e significativo per la popolazione Serba (STUD) e valori negativi e significativi per le popolazioni distribuite nell'areale occidentale della specie (ASPR e VALD).

Figura 4.12: Distribuzione dei valori del test D di Tajima tra le sette popolazioni. La significatività del test è indicata da * ($P < 5\%$) e da ** ($P < 1\%$).



I test di Fu & Li (D^* e F^* -test), mostrati in figura 4.13 e 4.14 rispettivamente, presentano entrambi un trend di valori positivi e significativi, in accordo con i risultati ottenuti in tabella 4.2, e che rafforzano l'ipotesi di deficit di mutazioni recenti. Le uniche eccezioni sono rappresentate dai geni 0_10384_02 e CL4470Ct1_01 che presentano valori negativi e significativi nella popolazione Corsa (VALD), compatibile con un'ipotesi di mutazioni recenti che hanno influenzato la distribuzione della diversità nucleotidica. Il gene 0_10384_02 mostra, per entrambi i test, un pattern diverso in base alla distribuzione della specie: valori positivi sono stati stimati nelle popolazioni distribuite ad est (STUD, MUGL e PNP) mentre valori negativi sono esclusivi delle popolazioni occidentali (ASPR, VALD e PNS).

In linea generale, i valori dei test di neutralità tra le varie popolazioni evidenziano che la maggior parte dei geni discostano in maniera significativa dal modello neutrale.

Figura 4.13: Distribuzione dei valori del test D* di Fu & Li tra le sette popolazioni. La significatività del test è indicata da * (P < 5%) e da ** (P < 1%).

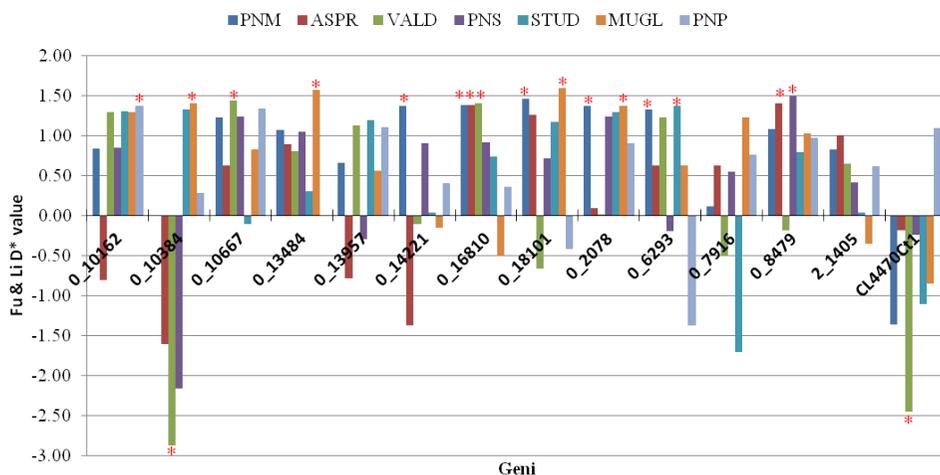
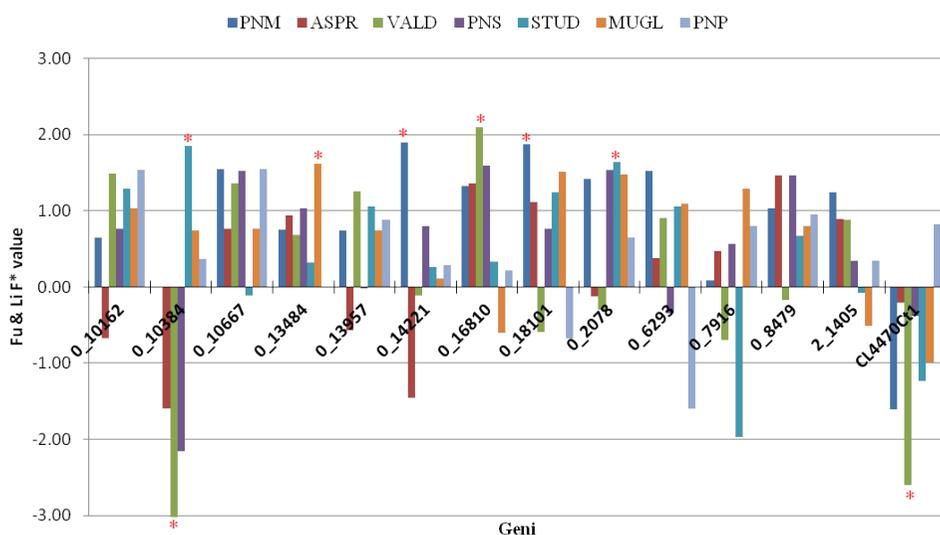
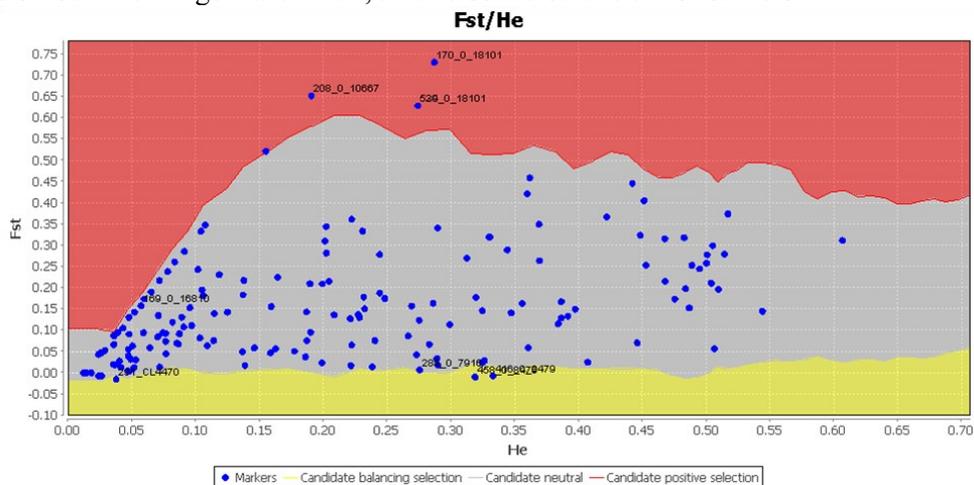


Figura 4.14: Distribuzione dei valori del test F* di Fu & Li tra le sette popolazioni. La significatività del test è indicata da * (P < 5%) e da ** (P < 1%).



I risultati dell'analisi degli outliers sono riportati nella figura 4.15, che fanno riferimento alle simulazioni effettuate considerando i 270 loci polimorfici (SNP e indel) dei 14 geni. I punti blu rappresentano i marcatori, le aree rossa e gialla indicano selezione positiva e selezione stabilizzante, rispettivamente.

Figura 4.15: Analisi dei loci candidati sotto selezione, tra i 270 siti polimorfici distribuiti nei 14 geni analizzati, attraverso metodiche di Fst-outliers.



Dopo la correzione dovuta al confronto tra ipotesi multiple (FDR, False Discovery Rate) sono stati individuati cinque loci potenzialmente sotto selezione. Tre loci presentano valori estremamente elevati di Fst, e ricadono nell'area che indica selezione positiva, mentre due loci presentano valori vicini allo zero e ricadono nell'area che indica selezione stabilizzante (Tabella 4.3).

Tabella 4.3: Loci sotto ipotetica selezione positiva o negativa sulla base dei risultati ottenuti dalle analisi degli Fst-outliers.

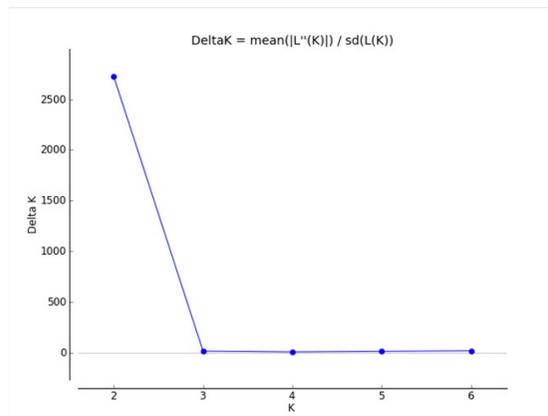
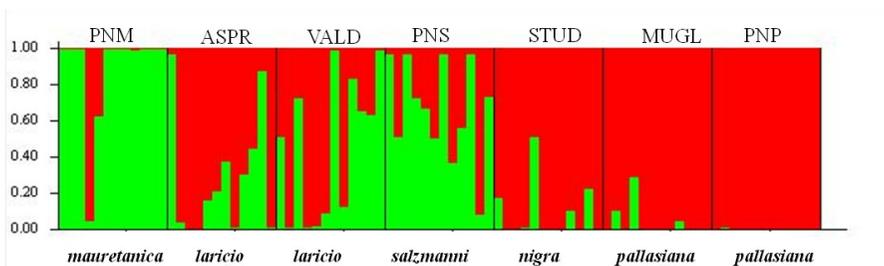
Selezione positiva	Selezione stabilizzante
208-0_10667_02	458-0_8479_01
169-0_16810_02	231-CL4470Ct1_01
170-0_18101_02	

4.2.3 Indagine molecolare a livello delle sottospecie.

La struttura genetica delle popolazioni è stata indagata attraverso metodi di raggruppamento Bayesiano. I cinque loci sotto ipotetica selezione sono stati esclusi per rispettare la condizione di neutralità richieste in questa serie di analisi, utilizzando quindi 265 loci SNP.

I risultati delle analisi effettuate con il programma STRUCTURE sono riportate in figura 4.16 e indicano la presenza di due gruppi principali.

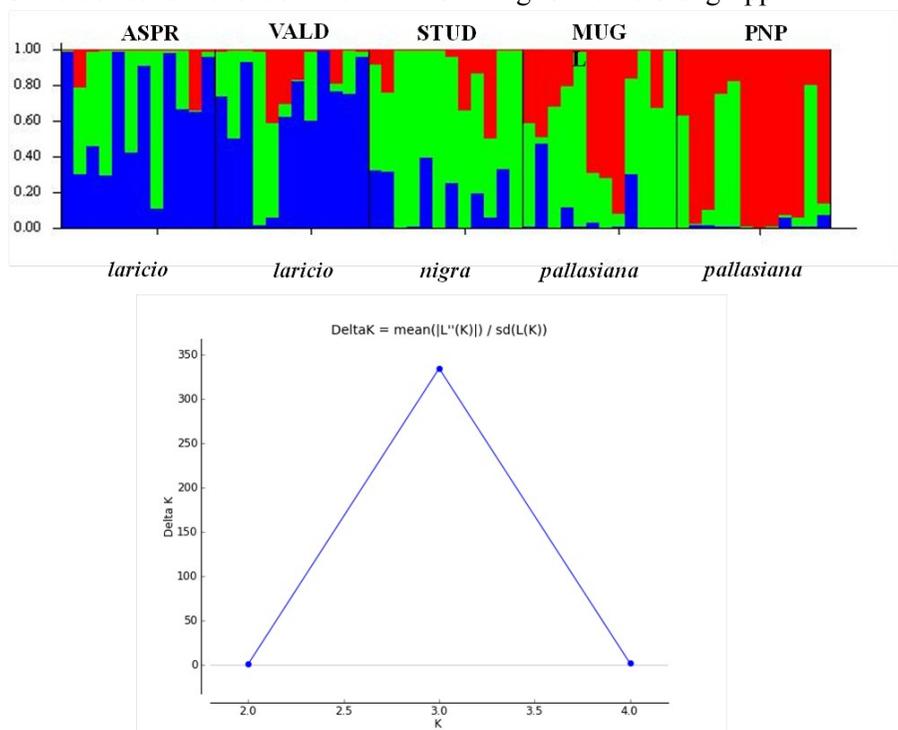
Figura 4.16: Analisi della struttura genetica delle popolazioni realizzata attraverso metodiche Bayesiane. Ogni istogramma rappresenta i 12 individui di ogni popolazione e ogni colore identifica la probabilità di appartenenza degli individui a quel gruppo. I risultati dell'analisi del deltaK che identifica in K=2 il miglior numero di gruppi è riportata nella parte inferiore della figura.



Un'ulteriore analisi Bayesiana è stata effettuata all'interno dei due gruppi in figura 4.16 per valutare la possibile presenza di sottostruttura. La figura 4.17 mostra i risultati ottenuti dall'analisi con STRUCTURE che evidenzia la presenza di sottostruttura, in particolare tre gruppi, uno comprendente le due

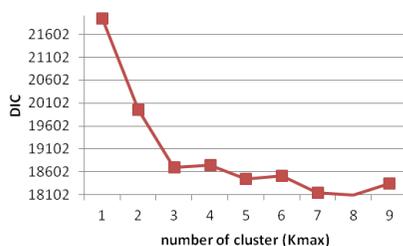
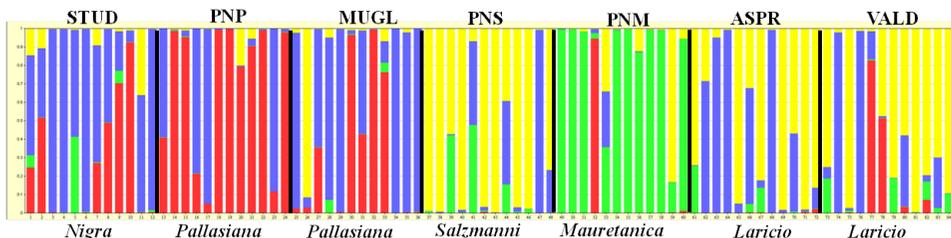
sottospecie laricio (ASPR e VALD), un altro le popolazioni Serba e Turca (STUD e MUGL), e il terzo la popolazione cipriota (PNP), appartenente alla sottospecie *pallasiana*.

Figura 4.17: Stima della sottostruttura in cinque popolazioni. Sotto è mostrata la stima del deltaK che identifica in K=3 il miglior numero di gruppi.



Per confermare la presenza di questa struttura genetica è stato utilizzato un altro approccio di *clustering* Bayesiano. La Figura 4.18 mostra i risultati dell'analisi di "clustering" con il software TESS. In questo caso il numero di gruppi identificati è uguale a quattro, con le tre popolazioni dell'areale orientale (STUD, PNP e MUGL) raggruppate in un cluster e le quattro popolazioni dell'areale occidentale che formano tre clusters, uno che comprende la popolazione Algerina (PNM), l'altro la popolazione francese (PNS) e l'ultimo le rimanenti due popolazioni (ASPR e VALD). In sostanza, I risultati di TESS confermano quelli di STRUCTURE.

Figura 4.18: Analisi della struttura delle popolazioni attraverso metodiche di classificazione Bayesiana TESS. Ogni istogramma rappresenta un individuo e ogni colore identifica la probabilità di appartenenza degli individui a quel gruppo. La stima del miglior numero di gruppi, identificata in $K=4$, è riportata sotto.



TESS è stato utilizzato per identificare possibili sottostrutture. La figura 4.19 mostra i risultati delle analisi effettuate sulle stesse popolazioni della figura 4.17.

I risultati riportati in figura 4.19 sono in accordo con quelli ottenuti con STRUCTURE. Sono stati identificati gli stessi gruppi riportati in figura 4.17, ma con una più elevata probabilità di appartenenza a quel determinato gruppo. TESS è stato anche utilizzato per valutare la struttura delle popolazioni occidentali e orientali: i risultati sono mostrati nelle figure 4.20 e 4.21, rispettivamente.

Figura 4.19: Analisi della sottostruttura in cinque popolazioni. La stima del miglior numero di gruppi, identificata in $K=3$, è riportata sotto.

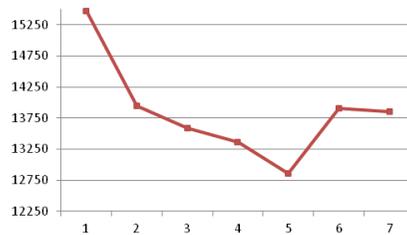
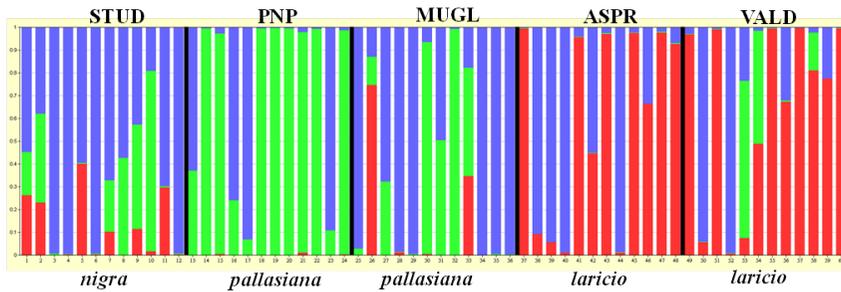


Figura 4.20: Risultati delle analisi di clustering Bayesiana utilizzando marcatori SNPs nelle popolazioni distribuite nell'areale occidentale della specie. La stima del miglior numero di gruppi, identificata in $K=3$, è riportata di sotto.

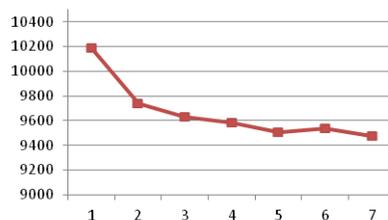
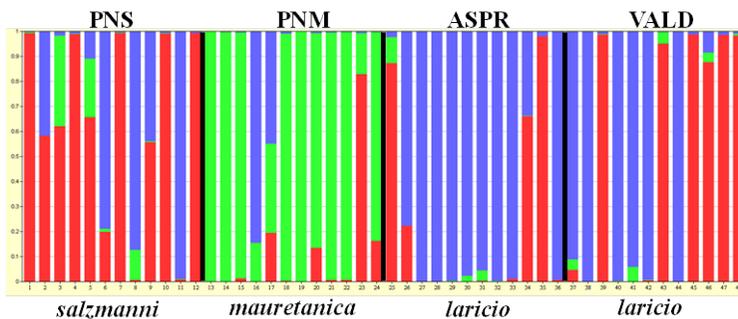
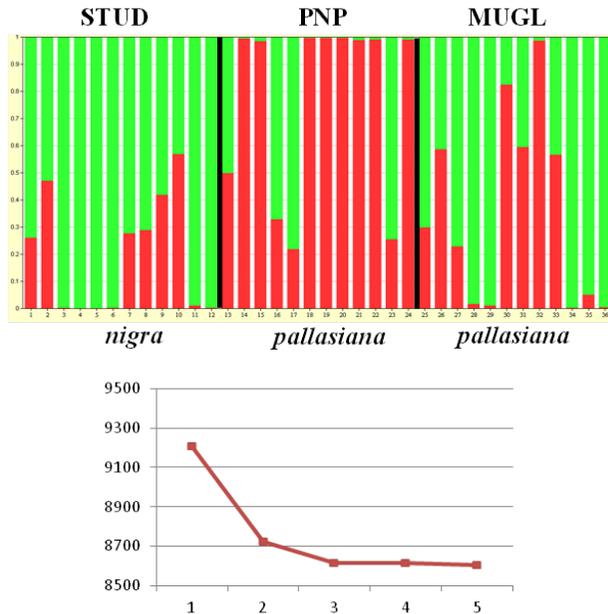


Figura 4.21: Analisi di clustering Bayesiana utilizzando marcatori SNPs sulle popolazioni distribuite nell'areale orientale della specie. Sotto è mostrata la stima del miglior numero di gruppi identificata in $K=2$.



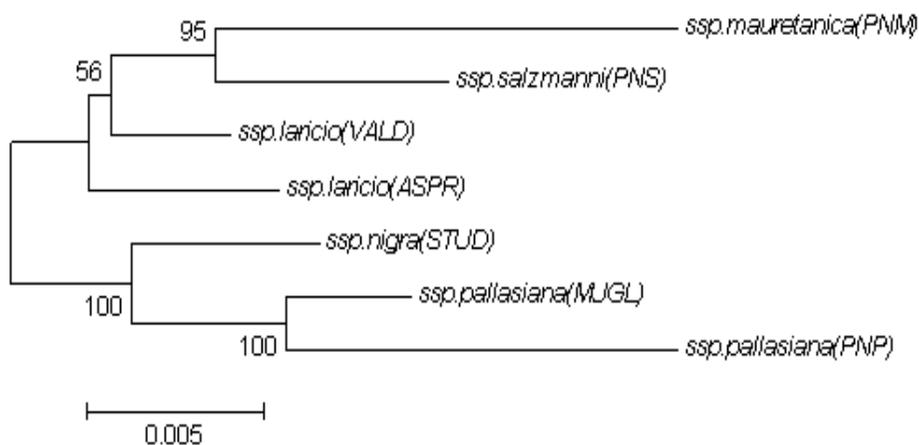
La tabella 4.4 riassume i valori di F_{st} stimati tra i vari gruppi identificati con le analisi di clustering Bayesiane.

Tabella 4.4: Differenziazione genetica tra i gruppi identificati con le analisi Bayesiane. Gli indici di fissazione (F_{st}) sono stati calcolati con il *software* Genetix.

Indici di fissazione tra gruppi	F_{st}
Fst all pop	0,0232
Fst ovest pop (PNM+ASPR+VALD+PNS)	0,0189
Fst est pop (STUD+MUGL+PNP)	0,0154
Fst_5 pop (ASPR+VALD+STUD+MUGL+PNP)	0,0169
Fst ssp laricio (VALD+ASPR)	0,0066
Fst ssp pallasiana (MUGL+PNP)	0,0143
Fst (MUGL+STUD)	0,0087

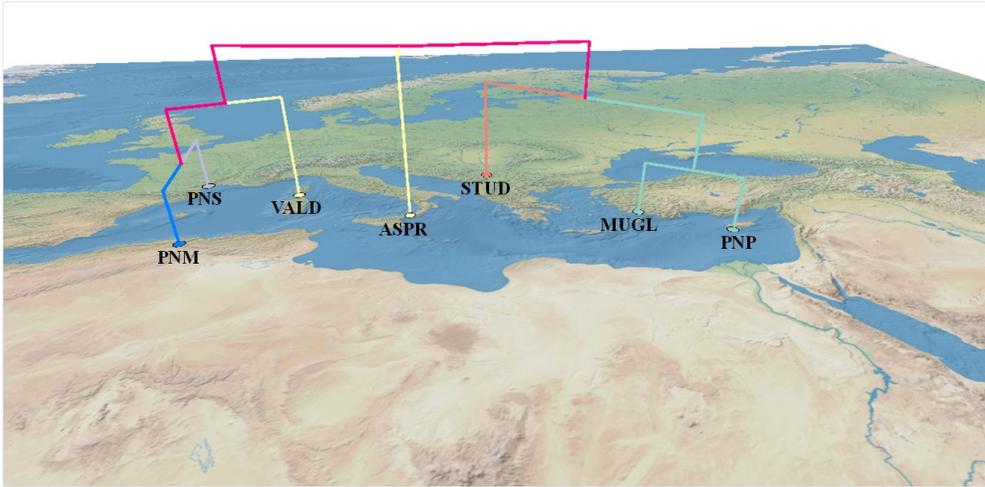
Sono stati utilizzati altri metodi di raggruppamento, per confermare o meno i risultati ottenuti con gli approcci Bayesiani. In particolare è stato disegnato un dendrogramma con il metodo *neighbour joint* (NJ) sulla matrice di distanze genetiche. I risultati sono mostrati in figura 4.22.

Figura 4.22: Dendrogramma *neighbour joint* (NJ) delle sette popolazioni di *Pinus nigra* utilizzando le distanze genetiche di Nei (D_A) calcolate su 265 loci SNP. I valori di *bootstrap* indicati dai numeri sui nodi del dendrogramma sono stati calcolati attraverso 1000 repliche.



La topologia del cluster mette in evidenza due gruppi principali in accordo con la distribuzione ovest-est delle popolazioni. La figura 4.23 mostra il dendrogramma NJ sovrapposto alla mappa di distribuzione delle popolazioni di pino nero analizzate.

Figura 4.23: Dendrogramma NJ sovrapposto alla mappa di distribuzione delle popolazioni analizzate.



All'interno del gruppo orientale, che comprende le popolazioni Serba (STUD), Turca (MUGL) e Cipriota (PNP), le due popolazioni appartenenti alla sottospecie pallasiana si raggruppano indicando un'elevata similarità genetica. Le popolazioni occidentali, anche se con meno supporto statistico rispetto alle popolazioni orientali, formano tre gruppi secondo le sottospecie di appartenenza e in accordo con le analisi Bayesiane.

Per confermare i risultati ottenuti con i marcatori SNP, è stata valutata la struttura genetica delle popolazioni anche con marcatori microsatelliti, sia plastidiali che nucleari. La figura 4.24 mostra i risultati dell'analisi TESS su cinque loci microsatelliti plastidiali, mentre la figura 4.25 quelli ottenuti analizzando quattro loci microsatelliti nucleari.

Figura 4.24: Risultati delle analisi della struttura genetica stimata con cinque loci microsatelliti plastidiali. La stima del miglior numero di gruppi, identificata in $K=2$, è riportata in basso.

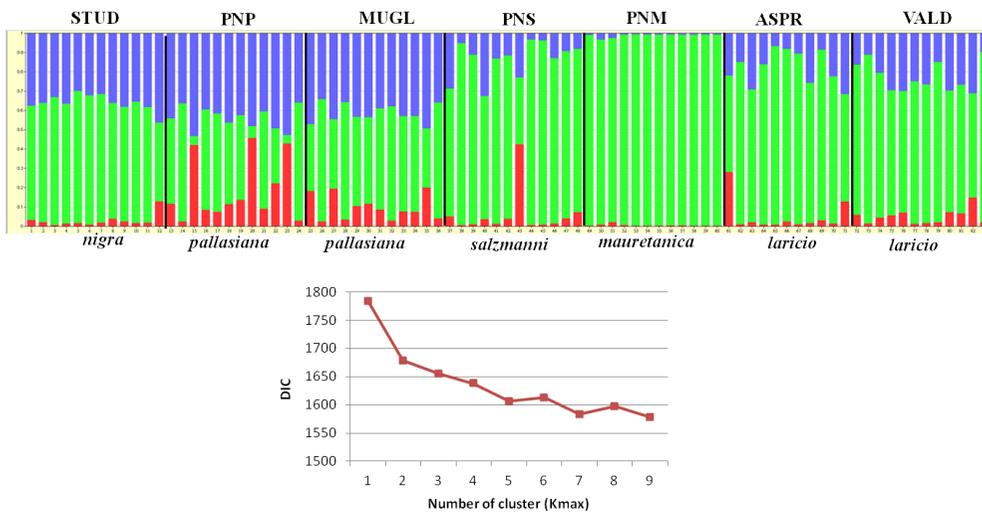
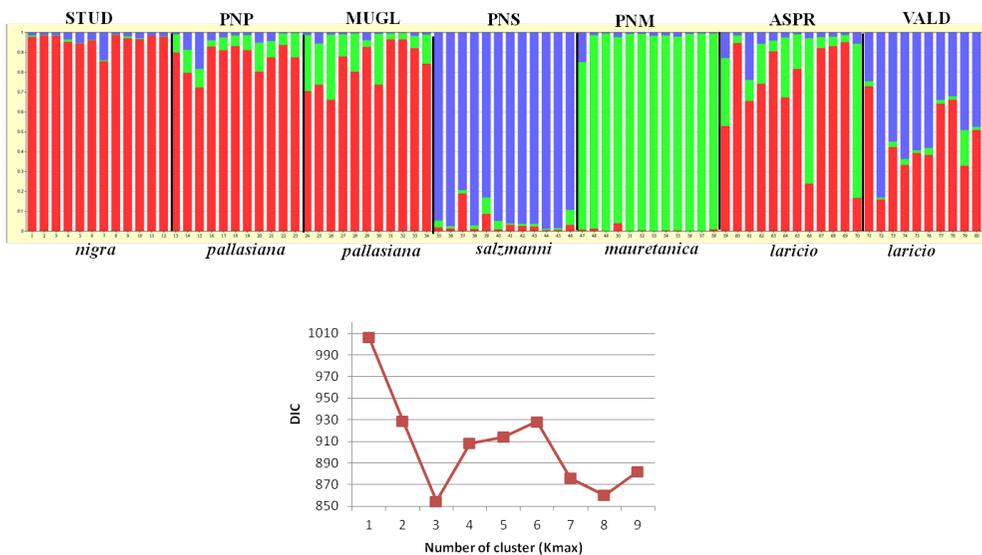


Figura 4.25: Risultati delle analisi della struttura genetica stimata con quattro loci microsatelliti nucleari. La stima del miglior numero di gruppi, identificata in $K=3$, è riportata di sotto.



Le analisi della struttura genetica utilizzando loci microsatelliti confermano, soprattutto per i loci nucleari, i risultati ottenuti dalle analisi con i marcatori SNP. I risultati ottenuti con i microsatelliti plastidiali mostrano una minore strutturazione tra le popolazioni, in particolare la popolazione Nord Africana (PNM) è rappresentata da un solo aplotipo fissato in tutti gli individui. Per quanto concerne la struttura ottenuta con gli SSR nucleari, sono stati identificati, all'interno delle popolazioni occidentali (PNS, PNM, ASPR e VALD), tre clusters che corrispondono alle tre differenti sottospecie. Le sottospecie salzmanni e mauretana formano due gruppi ben separati, mentre le due popolazioni appartenenti alla sottospecie laricio hanno un elevato grado di mescolamento; infatti, la popolazione Corsa (VALD) condivide diversi alleli della popolazione Francese (PNS), mentre la popolazione del sud Italia (ASPR) condivide alleli con le popolazioni orientali. Le popolazioni orientali formano un gruppo abbastanza omogeneo (MUGL e PNP). Come fatto per i marcatori SNP, anche per quelli microsatellite le popolazioni occidentali (Figura 4.26) e quelle orientali (Figura 4.27) sono state analizzate separatamente. La figura 4.26 conferma che la popolazione Corsa (VALD), appartenente alla sottospecie laricio, è molto più simile alla popolazione francese (PNS) che non alla popolazione localizzata nel sud Italia (ASPR) identificata anche essa come sottospecie laricio.

I risultati relativi alle tre popolazioni orientali (Figura 4.27) evidenziano, a differenza di quanto ottenuto utilizzando i marcatori SNP, che le due popolazioni della sottospecie pallasiana (MUGL e PNP) si raggruppano in due gruppi distinti, con la popolazione Turca (MUGL) molto più simile a quella Serba (STUD), appartenente alla sottospecie nigra, che non alla popolazione cipriota (PNP), appartenente alla stessa sottospecie.

Figura 4.26: Risultati delle analisi di clustering Bayesiana realizzata con quattro loci microsatelliti nucleari sulle popolazioni distribuite nell'areale occidentale della specie. La stima del miglior numero di gruppi, identificata in $K=3$, è riportata sotto.

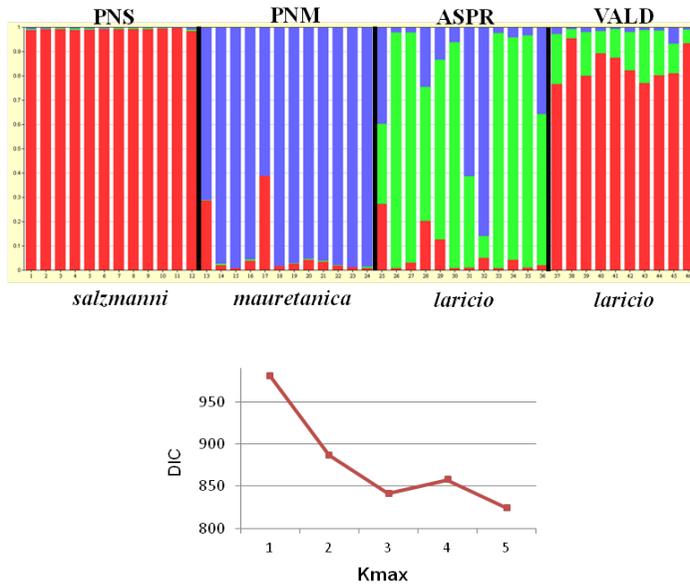
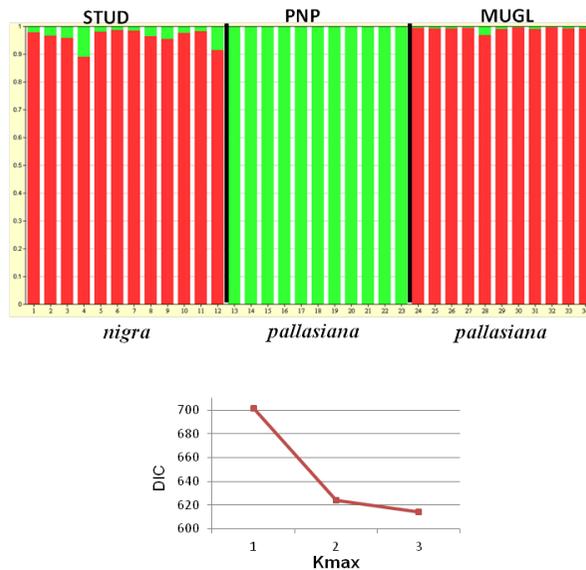


Figura 4.27: Risultati delle analisi di clustering Bayesiana ottenuta con quattro loci microsatelliti nucleari sulle popolazioni distribuite nell'areale orientale della specie. La stima del miglior numero di gruppi, identificata in $K=2$, è riportata sotto.



5. DISCUSSIONE

L'obiettivo principale di questo lavoro di tesi è stato quello di stimare la diversità genetica in popolazioni naturali di *Pinus nigra* attraverso l'utilizzo di marcatori SNP identificati e validati in 14 geni candidati. Questi geni sono stati precedentemente selezionati da un dataset di 236 sequenze geniche ipoteticamente coinvolte in vari processi legati all'adattamento (tolleranza/resistenza a stress abiotici e biotici e fenomeni legati alla fenologia) i cui primers sono stati disegnati in *Pinus taeda*, una specie di pino del Nord-America.

Queste 236 sequenze sono state editate manualmente, da un dataset iniziale di 264 geni, le cui sequenze di bassa qualità sono state eliminate, per una percentuale di trasferimento dell'89%, molto elevata se si pensa che le due specie appartengono a due sezioni differenti (sezione *Pinus*: *Pinus nigra* e sezione *Trifoliae*: *Pinus taeda*). Di questi 236 geni il 93% (219 sequenze geniche) è risultato polimorfico e solo il 7% monomorfico (11 sequenze geniche). La creazione di questo dataset di geni candidati ortologi di qualità elevata sarà utilizzato nell'ambito dell'iniziativa CRIEC, un progetto di risequenziamento nelle principali conifere Europee, per studi di genomica comparativa (Krutovsky et al., 2004), di associazione genetica (Eckert et al., 2009) e di genetica delle popolazioni (Cheddadi et al., 2006), che possono far luce su innumerevoli processi come la speciazione, la stima del potenziale adattativo delle specie e la comprensione del ruolo relativo dei processi demografici e selettivi sulla distribuzione della diversità nucleotidica.

In accordo con i primi studi sulle specie arboree forestali nei quali è stato osservato una frequenza elevata di SNP (Neale & Ingvarsson 2008; Savolainen & Pyhäjärvi 2007) anche nel nostro lavoro è stato ritrovato un elevato numero di polimorfismi di sequenza.

A livello delle sequenze geniche analizzate sono state identificate sia mutazioni puntiformi (SNP) che inserzioni-delezioni, per un totale di 270 polimorfismi, con una frequenza di circa 1 SNP ogni 100 bp; la frequenza osservata è dello stesso ordine di grandezza di quella riportata per altre specie appartenenti al genere *Pinus*, come *Pinus taeda* (González-Martínez et al., 2006) e *Pinus sylvestris* (Pyhäjärvi et al., 2007) e simile anche ad altre latifoglie forestali, come riportato ad esempio in uno studio recente sul

faggio (*Fagus sylvatica*) per il quale è stata calcolata una frequenza (in 10 geni candidati) di circa 1 SNP ogni 90 bp (Sieferet et al., 2012).

La diversità nucleotidica in *P. nigra* calcolata come valore medio sui 14 geni analizzati è risultata pari a $\pi = 0,0058$, valore analogo a quello ritrovato per altre specie appartenenti al sottogenere *Pinus*, con valori, stimati su un alto numero di geni candidati (in Mosca et al., 2012 hanno utilizzato oltre 800 sequenze di geni candidati), di π che variano da 0,0024 a 0,0082 (Ma et al., 2006; Eveno et al., 2008; Grivet et al., 2009; Wachowiak et al., 2009; Shiraishi & Shirishi 2011; Mosca et al., 2012) e in accordo con le stime di diversità nucleotidica ritrovate in altre specie montane appartenenti sempre al sottogenere *Pinus*, quali *P. sylvestris* (Pyhäjärvi et al., 2007; Palmé et al., 2008), *P. luchuensis* (Shiraishi & Shiraishi 2011) e *P. mugo* (Mosca et al., 2012). I livelli di diversità nucleotidica ritrovati in *Pinus nigra* sono concordi anche con quelli di altre specie di latifoglie come quelle appartenenti al genere pioppo ($\pi = 0,0072$ per *Populus nigra* in Chu et al. 2009; $\pi = 0,0042$ per *Populus tremula* in Ingvarsson 2008).

Le popolazioni appartenenti all'areale orientale di distribuzione della specie (STUD, MUGL e PNP, rispettivamente localizzate in Serbia, Turchia e Cipro) manifestano una maggiore diversità rispetto alle popolazioni occidentali (PNM, ASPR, VALD, PNS, rispettivamente localizzate in Algeria, sud Italia, Corsica e Francia) (Figura 4.8). La bassa diversità genetica della popolazione cipriota (PNP) può essere spiegata dalla ridotta dimensione della popolazione effettiva e dal ridotto flusso genico dovuto all'isolamento geografico di quest'isola. Stesso discorso è possibile fare per le popolazioni Calabrese (ASPR) e Corsa (VALD).

Questi risultati non possono essere corroborati da altri studi molecolari visto che quelli esistenti prendono in considerazione solo piccole parti dell'areale di distribuzione del pino nero (Nikolić & Tucić 1983; Tolun et al., 2000; Scaltsoyiannes et al., 2009; Jaramillo-Correa et al., 2010; Soto et al., 2010; Moraga et al., 2012). L'unico lavoro molecolare che comprende un numero consistente di popolazioni è quello di Afzal-Rafii & Dodd (2007) i quali hanno considerato la parte occidentale della distribuzione della specie (Marocco, Spagna, Sud-Italia, Corsica, Francia meridionale e Alpi orientali) ritrovando elevati livelli di diversità, stimata con marcatori microsatelliti cpDNA, nella maggior parte delle popolazioni analizzate, includendo quindi buona parte ma non tutte le entità tassonomiche di questa complessa specie.

Inoltre il presente lavoro al momento è l'unico che da una stima della diversità nucleotidica in questa specie.

L'analisi della varianza molecolare (Figura 4.10) mostra per tutti i geni valori significativi di differenziazione genetica (F_{st}); in media il 20% della variazione molecolare è dovuta a differenze tra popolazioni, con valori estremi che vanno da 35% (gene 0_16810_02) al 4% (gene 0_8479_01). Per valutare la quota di varianza dovuta a differenze tra le popolazioni, l'AMOVA è stata eseguita considerando due modelli: i) suddividendo le popolazioni in due gruppi che corrispondono alla distribuzione est-ovest dell'areale della specie; ii) suddividendo le popolazioni in cinque gruppi sulla base della loro appartenenza a sottospecie. Il modello a cinque gruppi, almeno sul set di geni considerati, spiega in maniera più significativa la quota di F_{sc} rispetto al modello a due gruppi. Il modello a due gruppi mantiene in media il 6% della variazione molecolare mentre per il modello a cinque gruppi oltre l'11% di variazione molecolare è dovuto a differenze tra le sottospecie (Figura 4.11).

I risultati sulla differenziazione tra le popolazioni mostrano inoltre che le popolazioni distribuite nell'areale occidentale sono leggermente più differenziate rispetto a quelle orientali (Tabella 4.4).

I risultati di diversità genetica e differenziazione sono in accordo con Scaltsoyiannes e collaboratori (2009) i quali affermano che questa specie non è stata influenzata in maniera significativa dagli eventi glaciali e potrebbe conservare ancora oggi una diversità genetica che già esisteva durante l'era Cenozoica, caratterizzata quindi da un'elevata differenziazione tra le popolazioni ed alto numero di alleli rari.

I risultati dei test di neutralità calcolati su tutta la popolazione hanno evidenziato che la maggior parte dei geni non discostano in maniera significativa dal modello neutrale. Il trend generale del test di Tajima mostra che 8 geni su 14 hanno valori negativi, indicando un eccesso di alleli rari, ma solo il gene CL4470Ct1_01 ha un valore significativamente differente da zero, il che può indicare una possibile selezione positiva che agisce su questo gene. I test di neutralità calcolati tra le popolazioni mostrano molti geni che discostano in maniera significativa dall'equilibrio neutrale. A differenza dei valori ritrovati quando si considerano tutti gli individui come appartenenti ad un'unica popolazione, si può notare come molti geni abbiano pattern contrastanti tra le popolazioni. In linea generale, i valori dei test di neutralità

tra le varie popolazioni evidenziano che la maggior parte dei geni discostano dal modello neutrale evidenziando probabilmente che il pattern della diversità nucleotidica è stato influenzato maggiormente dai fenomeni demografici, che agiscono sull'intero genoma, piuttosto che da fenomeni selettivi.

Questi risultati vanno comunque confermati con test di neutralità più potenti (come quelli composti) e con simulazioni Bayesiane approssimate (ABC test) che permettono di ricostruire la storia demografica passata e consentono di fare inferenze più robuste sulla storia evolutiva di questa specie.

A livello tassonomico è ancora molto dibattuta la questione della possibile presenza di entità sottospecifiche. In questo lavoro sono state verificate le possibili relazioni tra le popolazioni campionate, che ricoprono la maggior parte dell'areale di distribuzione Mediterraneo. La distribuzione della diversità nucleotidica sembra in parte essere riconducibile alla presenza di sotto-specie e confermare lo status di complex del *Pinus nigra*.

La popolazione Algerina, appartenente alla sottospecie mauretana, è molto probabile che si sia originata da popolazioni Spagnole (Bogunic et al., 2007; Jaramillo-Correa et al., 2010). Jaramillo-Correa e collaboratori (2010) riportano inoltre che le regioni del Nord-Africa (Marocco e Algeria) sono state colonizzate, oltre che da popolazioni provenienti dalla Spagna, anche dal Sud-Italia con intenso flusso genico via polline, che ha permesso di attraversare lo stretto di Sicilia e che ha dato origine a fenomeni di ibridazione ed introgresione con le popolazioni nord Africane. I risultati ottenuti con gli approcci Bayesiani hanno mostrato che le popolazioni Algerine e Italiane si raggruppano in due clusters ben separati, in accordo con le due sottospecie di appartenenza, mauretana e laricio rispettivamente. E' da sottolineare che questo è il primo studio che mette in luce le possibili affinità tra le popolazioni Africane e quelle del Sud-Italia utilizzando sia marcatori SNPs (Figura 4.20) che marcatori microsatelliti nucleari (Figura 4.26).

La popolazione Corsa (ssp. *laricio*) costituisce il cluster più controverso; infatti, indagini recenti basate sul confronto dei pattern di terpeni e cpSSRs (Bojovic et al., 2005; Afzal-Rafii & Dodd 2007) trovano una moderata similarità con le popolazioni Francesi (ssp. *salzmanni*). Inoltre gli scarsi ritrovamenti palinologici sull'isola prima dell'intensificazione delle attività antropiche potrebbero suggerire che queste popolazioni non sono autoctone

ma si sono originate con materiale proveniente dalla Francia e che queste si siano successivamente differenziate per isolamento (Reille et al., 1997). D'altro canto altri lavori hanno evidenziato una certa similarità tra le popolazioni Corse e quelle del Sud-Italia, tanto da ipotizzare una loro comune appartenenza alla ssp. *laricio*. Afzal-Rafii & Dodd (2007), sulla base di simulazioni demografiche, ottengono risultati che permettono di avanzare un'ipotesi alternativa che prevede la presenza della specie sull'isola durante l'ultima glaciazione e stima una dimensione effettiva della popolazione molto piccola (1600-3600 individui), che spiegherebbe la scarsità dei ritrovamenti palinologici. I nostri risultati, ottenuti con metodiche Bayesiane, evidenziano che le popolazioni Corse, del Sud-Italia e della Francia meridionale a livello dei marcatori SNPs (Figura 4.20) e SSRs (Figura 4.26) mostrano elevati livelli di mescolamento non consentendoci di escludere nessuna delle due ipotesi sopra elencate. Per quanto concerne le popolazioni distribuite nell'areale orientale della specie (Balcani e Turchia), le analisi di clustering con i marcatori SNPs (Figura 4.21) mostrano alti livelli di mescolamento, mentre i risultati con i marcatori SSR nucleari (Figura 4.27) evidenziano che le due popolazioni della sottospecie *pallasiana* (MUGL e PNP) si raggruppano in due gruppi distinti, con la popolazione Turca (MUGL) molto più simile a quella Serba (STUD), appartenente alla sottospecie *nigra* o *dalmatica*, che non alla popolazione cipriota (PNP), appartenente alla stessa sottospecie. A causa dell'assenza di popolazioni dell'arco Alpino e della Croazia non possiamo valutare una possibile appartenenza della popolazione Serba alla ssp. *nigra* o ssp. *dalmatica*. Inoltre considerando che queste zone costituiscono la maggior parte dell'areale della specie e che la sua diffusione in Europa è probabilmente iniziata a partire da queste regioni, un campionamento più dettagliato sarebbe necessario. Questo lavoro comunque è il primo che utilizza marcatori molecolari nucleari per valutare le relazioni tra popolazioni occidentali e orientali.

La strategia di SNP-discovery in geni candidati, utilizzata in questo lavoro, è al momento la migliore in specie appartenenti alla famiglia delle Pinaceae dove la mancanza di un genoma di riferimento e le dimensioni enormi dei genomi stessi non consentono la scelta di altre strategie (ad esempio il re-sequenziamento di interi genomi) per monitorare la diversità genetica a livello di intero genoma (wide genome scan).

Ringraziamenti

Il presente lavoro di tesi è stato svolto nell'ambito di un progetto internazionale di risequenziamento di un elevato numero di geni candidati coinvolti in processi relativi all'adattamento agli stress abiotici nelle principali conifere Europee (progetto CRIEC, Conifer Re-sequencing Initiative European Community) coordinato dall'Istituto di Genetica Vegetale del CNR di Firenze.

Riferimenti bibliografici

Achtak H, Oukabli A, Ater M, Santoni S, Kjellberg F, Khadari B (2009). Microsatellite Marker as Reliable Tool for Fig Cultivar Identification. *Journal of American Society Hort Sci* 134(6):624-631.

Adawy SS, Hussein EHA, Ismail SEME, El-Itriby HA (2006) Genomic diversity in date palm (*Phoenix dactylifera* L.) as revealed by AFLPs in comparison to RAPDs and ISSRs. *Arab. Jour. Biotech* 8: 99-114.

Afzal-Rafii Z & Dodd RS (2007). Chloroplast DNA supports a hypothesis of glacial refugia over postglacial recolonization in disjunct populations of black pine (*Pinus nigra*) in western Europe. *Molecular Ecology*, 16: 723-36.

Aguinagalde I & Llorente F (1997). Relationships among five populations of European black pine (*Pinus nigra* Arn.) using morphometric and isozyme markers. *Sil Gen.* 46: 1-5.

Ahmed TA & Al-Qaradawi AY (2009). Molecular phylogeny of Qatari date palm genotypes using Simple Sequence Repeats markers. *Biotech* 8: 126-131.

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. (2008). Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* 1: 95-111.

Akkak A, Scariot V, Torello Marinoni D, Boccacci P, Beltramo C, Botta R (2009). Development and evaluation of microsatellite markers in *Phoenix dactylifera* L. and their transferability to other Phoenix species. *Biologia Plantarum* 53(1): 164-166.

Al-Dous EK, George B, Al-Mahmoud ME et al (2011) De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nature Biotechnology* 29(6): 521–527.

Al-Farsi MA & Lee CY (2008). Nutritional and functional properties of dates: a review. *Crit. Rev. Food Science Nutrition* 48: 877-887.

Al-Mamhoud ME, Al-Dous EK, Al-Azwani EK, Malek JA (2012). DNA-based assays to distinguish date palm (*Aracaceae*) gender. *American Journal of Botany* e7-e10.

Al-Ruqaishi IA, Davey M, Alderson P, Mayes S (2008) Genetic relationships and genotype tracing in datepalm (*Phoenix dactylifera* L.) in Oman based on microsatellite markers. *Genet Resour Crop Evol* 61: 70-72.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008). LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* 9: 323-328.

Aradhya MK, Dangel GS, Prins BH, Boursiquot JM, Walker MA, Meredith CP, Simon CJ (2003). Genetics structure and differentiation in cultivated grape, *Vitis vinifera* L. *Genetical Research* 81: 179-192.

Arumuganathan K & Earle ED (1991). Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9: 208-218.

Asmussen CB & Chase MW (2001). Coding and noncoding plastid DNA in palm systematics. *American Journal of Botany* 88: 1103-1117.

Barrett HC (1973). Date breeding and improvement in North America. *Fruit Varieties Journal* 27: 50-55.

Awadalla P & Ritland K (1997) Microsatellite variation and evolution in the *Mimulus guttatus* species complex with contrasting mating systems. *Mol Biol Evol* 14: 1023-1034.

Barrow S (1998). A monograph of *Phoenix* L. (Palmae: Coryphoideae). *Kew Bulletin* 53: 513-575.

Beal JM (1937). Cytological studies in the genus *Phoenix*. *Botanical Gazette* 99: 400-407.

Bekheet SA (2011). In vitro conservation of date palm germplasm. In: *Date Palm Biotechnology*, Springer, Netherland, Pp 337-360.

Belkhir K, Goudet J, Chikhi L. et al., (2000). Genetix (Ver. 4.01), logiciel sous windows™ pour la géne'tique des populations. Laboratoire Ge'nome et Population, Université' Montpellier II, Montpellier, France.

Besnard G & Bervillè A (2000). Multiple origins for Mediterranean olive (*Olea europaea* L. *ssp. europaea*) based upon mitochondrial DNA polymorphism. *Comptes Rendus de L'Académie des Sciences de la Vie, Life Sciences* 323: 173-181.

Billotte N, Marseillac N, Brottier P, Noyer JL, Jacquemoud-Collet JP, Moreau C, Couvreur T, Chevallier MH, Pintaud JC, & Risterucci AM (2004). Nuclear microsatellite markers for the date palm (*Phoenix dactylifera* L.): characterization and utility across the genus *Phoenix* and in other palm genera. *Molecular Ecology Notes* 4: 256-258.

Biswas S & Akey JM (2006). Genomic insights into positive selection. *Trends Genet* 22: 437-446.

Bodiain A, Ndir KN, Diedhiou PM, Borgel A, Sane D, Sagna M, Chevalier MH (2012a). Analysis of genetic diversity of date palm (*Phoenix dactylifera* L.) cultivars from Mauretania using microsatellites markers. *International Journal of Science and Advanced Technology* 2: 36-43.

Bodiain A, El Houmaizi MA, Ndir KN, Hasnaoui A, Nachtigall M, Wehling P (2012b). Genetic diversity analysis of date palm (*Phoenix dactylifera* L.) cultivars from Figuig (Morocco) using SSR markers. *International Journal of Science and Advanced Technology* 2: 96-104.

Bogunic F, Muratovic E, Ballian D, Siljak-Yakovlev S, Brown S (2007). Genome size stability among five subspecies of *Pinus nigra* Arnold s.l. *Environ Exp Bot* 59: 354-360.

Bojovic S, Jurc M, Drazic D, Pavlovic P, Mitrovic M, Djurdjevic L, Dodd RS, Afzal-Rafii Z, Barbero M (2005). Origin identification of *Pinus nigra* populations in southwestern Europe using terpene composition variations. *Trees* 19: 531-538.

Bonan GB (2008). Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* 320 : 1444-1449.

Botta R, Scott NS, Eynard I, Thomas MR (1995). Evaluation of microsatellite sequence tag site markers for characterizing *Vitis vinifera* cultivars. *Vitis* 34: 99-102.

Bourgis F, Kilaru A, Cao X et al (2011). Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proc Nat Acad Sci USA* 108(30): 12527-12532.

Brac de la Perriere RA & Benkhalifa A (1991). Progression de la du palmier dattier en Algérie. *Sécheresse* 2: 119-128.

Bracci T, Sebastiani L, Busconi M, Fogher C, Belaj A, Trujillo I (2009). SSR markers reveal the uniqueness of olive cultivars from the Italian region of Liguria. *Scientia Horticulturae* 122: 209-215.

Breton C, Tersac M, Bervillè A (2006). Genetic diversity and gene flow between the wild olive (oleaster, *Olea europaea* L.) and the olive: Several Plio-Pleistocene refuge zone in the Mediterranean basin suggested by simple sequence repeat analysis. *Journal of Biogeography* 33: 1916-1928.

Bronzini de Caraffa V, Maury J, Gambotti C, Breton C, Bervillè A, Giannettini J (2002). Mitochondrial DNA variation and RAPD mark oleaster, olive and feral olive from Eastern and Western Mediterranean. *Theoretical and Applied Genetics* 104: 1209-1216.

Brown S, Sathaye J, Cannell M, Kauppi PE (1996). Mitigation of carbon emissions to the atmosphere by forest management. *J Comm Forestry Review* 75(1): 80-91.

Brown GR, Kadel III EE, Bassoni DL, Kiehne KL, Temesgen B, van Buijtenen JP, Sewell MM, Marshall KA, Neale DB (2001). Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. *Genetics* 159: 799-809.

Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004). Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proc Natl Acad Sci USA* 101: 15255-15260.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol Evol* 18: 249–256.

Bucci G, Anzidei M, Madaghiele A, Vendramin GG (1998). Detection of haplotypic variation and natural hybridization in halepensis-complex pine species using chloroplast simple sequence repeat (SSR) markers. *Mol Ecol* 7: 1633–1643.

Camarero JJ, Gutierrez E, Fortin MJ, Ribbens E (2005). Spatial patterns of tree recruitment in a relict population of *Pinus uncinata*: forest expansion through stratified diffusion. *J Biogeography* 32(11): 1979-1992.

Cao BR & Chao CT (2002). Identification of date cultivar in California using AFLP markers. *HortScience* 37: 966-968.

Chambel MR, Climent J, Alia R (2007). Divergence among species and populations of Mediterranean pines in biomass allocation of seedlings grown under two watering regimes. *Ann For Sci* 64: 87-97.

Chao CT & Krueger RR (2007). The date palm (*Phoenix dactylifera* L.): overview of biology, uses and cultivation. *HortScience* 42: 1077-1082.

Charlesworth B, Sniegowski P, Stephan W (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371: 215-220.

Cheddadi R, Vendramin GG, Litt T, François L, Kageyama M, Lorentz S, Laurent JM, De Beaulieu JL, Sadori L, Jost A, Lunt D (2006). Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. *Global Ecology and Biogeography*, 15: 271-282.

Chen C, Durand E, Forbes F, François O (2007). Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. *Mol Ecol Notes* 7: 747-756.

Chu Y, Su X, Huang Q, Zhang X (2009). Patterns of DNA sequence variation at candidate gene loci in black poplar (*Populus nigra* L.) as revealed by single nucleotide polymorphisms. *Genetica* 137: 141-150.

Chung SM & Staub JE (2003). The development and evaluation of consensus chloroplast primer pairs that possess highly variable sequence regions in a diverse array of plant taxa. *Theor Appl Genet* 107: 757-767.

Climent J, Prada MA, Calama R, Chambel MR, de Ron DS, Alía R (2008). To growth or to seed: ecotypic variation in reproductive allocation and cone production by young female Aleppo pine (*Pinus halepensis*) Pinaceae. *Am J Bot* 95: 1-10.

Coart E, Vekemans X, Smulders MJM, Wagner I, Van Huylenbroeck J, Van Vockstaele E, Roland-Ruiz I (2003). Genetic variation in the endangered wild apple (*Malus sylvestris* L, Mill) in Belgium as revealed by AFLP and SSRs markers. *Molecular Ecology* 12: 845-857.

Cozzolino S, Cafasso D, Pellegrino G, Musacchio A, Widmer A (2003). Molecular evolution of a plastid tandem repeat locus in an Orchid lineage. *Journal of Molecular Evolution* 57: 41-49.

Cronn RC, Small RL, Haselkorn T, Wendel JF (2002). Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *American Journal of Botany* 89: 707-725.

Cronn R, Liston A, Parks M, Gernandt DS, Shen R, Mockler T (2008) Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucl Acids Res* 36: e122.

Dangl GS, Mendum ML, Prins BH, Walker MA, Meredith CP, Simon CJ (2001). Simple sequence repeat analysis of a clonally propagated species: A tool for managing a grape germplasm collection. *Genome* 44: 432-438.

Dansgaard W, Johnsen SJ, Clausen HB, Dahl-Jensen D, Gundestrup NS, Hammer CU, Hvidberg CS, Steffensen JP, Sveinbjornsdottir AE, Jouzel J, Bond G (1993). Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* 364: 218-220.

Daskalakou EN & Thanos CA (1996). Aleppo pine (*P. halepensis*) Postfire Regeneration: the role of canopy and soil seed banks. *International Journal of Wildland Fire* 6(2): 29-66.

De Cillis E (1923). Saggio di “Fenicigrafia Libica” – Studi sopra alcune razze di palma da datteri coltivate in Tripolitania. *Bollettino di informazione economica* n.6. Roma, Italia.

Demesure B, Sodji N, & Petit RJ (1995). A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molec. Ecol.* 4: 129–131.

Deulvot C, Charrel H, Amandine M, Jaquin F, Donnadieu C, Hennaut IL, Burstin J, Aubert G (2010). Highly multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. *BMC Genomics* 11: 468-478.

Djerbi M (1982). Bayoud disease in North Africa: history, distribution, diagnosis and control. *Date Palm Journal* 1: 153-197.

Dodd ME & Silvetown J (2000). Size-specific fecundity and the influence of life time size variation upon effective population size in *Abies balsamea*. *Heredity* 85: 604-609.

Doebley JF, Gaut BS, Smith BD (2006). The molecular genetics of crop domestication. *Cell* 127: 1309-1321.

Doyle JJ, Morgante M, Tingey SV, Powell W (1998). Size homoplasy in chloroplast microsatellites of wild perennial relatives of soybean (*Glycine subgenus Glycine*). *Mol Biol Evol* 15: 215-218.

Doveri S, Lee D, Maheswaran M, Powell W (2008). Molecular markers: History, features and applications. In: Principles and practices of plant genomics, Vol I, Abbott ed. (Enfield, USA Science Publishers). pp. 23-68.

Dransfield J, Uhl NW, Asmussen CB, Baker WJ, Harley MM, Lewis CE (2008a). Genera palmarum: The evolution and classification of palms. Royal Botanic Gardens, Kew, UK.

Dransfield J, Rakotoarinivo M., Baker WJ, Bayton RP, Fisher JB, Horn J, Leroy B, & Metz X (2008b). A new Coryphoid palm genus from Madagascar. *Botanical Journal of the Linnean Society* 156: 79-91.

Duran C, Appleby N, Edwards D, Batley J (2009). Molecular Genetic Markers: Discovery, Applications, Data Storage and Visualisation. *Curr Bioinformatics* 4: 16-27.

Durand E, Jay F, Gaggiotti OE, François O (2009). Spatial inference of admixture proportions and secondary contact zone. *Mol Biol Evol* 26 :1963-1973.

Earl DA & vonHoldt BM (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4 (2): 359-361.

Edgar RC (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 19:1792-1797.

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996). Microsatellite library en-riched for several microsatellite sequence in plant. *BioTechniques* 20: 758-760.

Ehrich D (2006). AFLPdat: a collection of R functions for convenient handling of AFLP data. *Mol Ecol Notes* 6: 603-604.

Eckert AJ & Hall BD (2006). Phylogeny, historical biogeography, and patterns of diversification for *Pinus* (Pinaceae): phylogenetic tests of fossil-based hypotheses. *Mol Phylogenet Evol* 40: 166-182.

Eckert AJ, Bower AD, Wegrzyn JL, Pande B, Jermstad KD, et al. (2009). Association genetics of coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-hardiness related traits. *Genetics* 182: 1289-1302.

Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB (2010). Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Mol Ecol* 19: 3789-3805.

El-Assar AM, Krueger RR, Devanands PS, Chao CT (2003). Genetic analysis of date palms (*Phoenix dactylifera* L.) from Egypt using fluorescent-AFLP markers. *Hort Sci* 38: 733-734.

El-Assar AM, Krueger RR, Devanands PS, Chao CT (2005). Genetic analysis of Egyptian date (*Phoenix dactylifera* L.) accession using AFLP markers. *Genet Resour Crop Evol* 52: 601-607.

Ellegren H (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* 12: 551-558.

Ellegren H (2004). Microsatellites: simple sequence with complex evolution. *Nature* 5: 435-446.

El Hadrami I & Baaziz M (1995). Somatic embryogenesis and analysis of peroxidase in *Phoenix dactylifera* L. *Biol Plant* 37: 197-203.

Elhoumaizi MA, Saaidi M, Oihabi A, Cilas C (2002). Phenotypic diversity of date palm cultivar (*Phoenix dactylifera* L.) from Morocco. *Genetic Resource Crop Evol* 49: 483-490.

Eke CR, Akomeah P, Asemota O (2005). Somatic embryogenesis in date palm (*Phoenix dactylifera*) from apical meristem tissues from “zebia” and “loko” landraces. *African Journal of Biotechnology* 4: 244-246.

Elmeer K, Baum M, Sarwath H, Hamwiah A, Malek J (2011). New microsatellite markers for assessment of genetic diversity in date palm (*Phoenix dactylifera* L.). *Biotech* 1: 91-97.

Elmeer K & Mattat I (2012). Marker assisted sex differentiation in date palm using simple sequence repeats. *3 Biotech* 2: 241-247.

Elshibli S & Korpelainen H (2008). Microsatellite markers reveals high genetic diversity in date palm (*Phoenix dactylifera* L.) germplasm from Sudan. *Genet* 134:251-260

Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulations study. *Molecular Ecology* 14: 2611-2620.

Eveno E, Collada C, Guevara MA, Léger V, Soto A, Díaz L, Léger P, González-Martínez SC, Cervera MT, Plomion C et al. (2008). Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Mol Biol Evol* 25: 417-437.

Ewing B, Hillier L, Wendl MC, Green P (1998). Base-calling of automated sequencer trace using phred. I. Accuracy assessment. *Genome Res* 8: 175-185.

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.

Excoffier L (2004). Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Mol Ecol* 13(4): 853-864.

Excoffier L & Lischer HEL (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* 10: 564-567.

Eyre-Walker A & Gaut BS (1997). Correlated rates of synonymous site evolution across plant genomes. *Molecular Biology and Evolution* 14:455-460.

Fady B (2005). Is there really more biodiversity in Mediterranean forest ecosystems? *Taxon* 54: 905–910.

Fady B & Conord C (2010). Macroecological patterns of species and genetic diversity in vascular plants of the Mediterranean Basin. *Divers Distrib* 16: 53–64.

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.

Fang Y, Wu H, Zhang T, Yang M, Yin Y, et al. (2012). A Complete Sequence and Transcriptomic Analyses of Date Palm (*Phoenix dactylifera* L.) Mitochondrial Genome. PLoS ONE 7(5): e37164.

FAO (1982). Plant Production and Protection: Paper No. 35. Date production and protection. FAO, Rome, Italy, pp: 294.

FAO (2010). Global Forest Resources Assessment 2010. Rome, pg 378.

Farjon A (2001). World checklist and bibliography of conifers, 2nd edn. The Royal Botanic Gardens, Kew, UK.

Fineschi S (1984). Determination of the origin of an isolated group of trees of *Pinus nigra* through enzyme gene markers. Silv Gen. 33: 169-172.

Flanary BE & Kletetschka G (2005). Analysis of telomere length and telomerase activity in tree species of various life-spans, and with age in the bristlecone pine *Pinus longeva*. Biogerontology 6(2): 101-111.

Frankham R (1997). Do island populations have less genetic variation than mainland populations? Heredity 78: 311-327.

Freeland JR (2005). Genetic analysis of single population (Chap 3). In: Molecular Ecology. Ed. Jonh Wiley and Sons Ltd, England. pp. 63-106.

Fu YX & Li WH (1993). Statistical tests of neutrality of mutations. Genetics 133: 693-709.

Fu YX (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147: 915-925.

Fukarek P (1958). Die Standortstrassen der Schwarzföhre (*Pinus nigra* Arn.). Centralblatt fuer das gesamte Forstwesen 75:203-207.

Fulton T, Van der Hoeven MR, Eannetta NT, Tanksley SD (2002). Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14: 1457-1467.

García-Gil MR, Mikkonen M, Savolainen O (2003). Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. Mol Ecol 12: 1195-1206.

Gaussen H (1960). Les Gymnospermes actuelles et fossils. Fascicule VI. Généralités, genre Pinus; Faculté des Sciences, Toulouse.

Gaut BS, Muse SV, Clark WD, Clegg MT (1992). Relative rates of nucleotide substitution at the *rbcL* locus in monocotyledonous plants. *Journal of Molecular Evolution* 35:292-303.

Gerber S, Baradat P, Marpeau A, Arbez M (1995). Geographic variation in terpene composition of *Pinus nigra* Arn. *Forest Gen* 2: 1–10.

Gernandt DS, Geada Lopez G, Ortiz Garcia S, Liston A (2005). Phylogeny and classification of *Pinus*. *Taxon* 54: 29-42.

Giorgi F & Lionello P (2008). Climate change projections for the Mediterranean region. *Glob Planet Change* 63: 90-104.

Glèmin S & Bataillon T (2009). A comparative view of the evolution of grasses under domestication. *New Phytologist* 183: 273-290.

Goldstein DB & Pollock DD (1997). Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Jour. Heredity* 88: 335-342.

Gómez A, Alía R, Bueno MA (2001). Genetic diversity of *Pinus halepensis* Mill. Populations detected by RAPD loci. *Ann Forest Sci* 58: 869–875.

González-Martínez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006). DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* 172: 1-12.

Gordon D, Desmarais C, Green P (2001). Automated finishing with Autofinish. *Genome Res* 11: 614-625.

Gore MA, Chia J-M, Elshire RJ, Sun Q, Ersoz ES, et al. (2009). A First-Generation Haplotype Map of Maize. *Science* 326: 1115-1117.

Goulao L & Oliveira CM (2001). Molecular characterisation of cultivars of apple (*Malus × domestica* Borkh.) using microsatellite (SSR and ISSR) markers. *122*: 81-89.

Govaerts R & Dransfield J (2005). World Checklist of Palms: 1-223. The Board of Trustees of the Royal Botanic Gardens, Kew.

Haddouch M (1996). Situation actuelle et perspectives de développement du palmier dattier au Maroc. *Options Mediterr.* 28: 63-79.

Grattapaglia D, Silva-Junior OB, Kirst M, de Lima BM, Faria DA, Pappas GJ (2011). High-throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species. *BMC Plant Biology* 11: 65-83.

Grivet D, Sebastiani F, González-Martínez SC, Vendramin GG (2009). Patterns of polymorphism resulting from long-range colonization in the Mediterranean conifer Aleppo pine. *New Phyt* 184: 1016-1028.

Grivet D, Sebastiani F, Alía R, Bataillon T, Torre S, Zabal-Aguirre M, Vendramin GG, González-Martínez SC (2011). Molecular footprints of local adaptation in two Mediterranean conifers. *Mol Biol Evol* 28: 101-116.

Grotkopp E, Rejmanek M, Sanderson MJ, Rost TL (2004). Evolution of genome size in pines (*Pinus*) and its life-history correlates: supertree analyses. *Evolution* 58: 1705-1729.

Hahn WJ (2002). A molecular phylogenetic study of the Palmae (Arecaceae) based on *atpB*, *rbcL*, and 18S nrDNA sequences. *Systematic Biology* 51: 92-112.

Hammadi H, Vendramin GG, Ali F (2011). Microsatellite diversity among Tunisian date palm (*Phoenix dactylifera* L.) subpopulations. *Pak J Bot* 43(2): 1257-1265.

Hamrick JL & Godt MJW (1990). Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding and Genetic Resource*. (Brown AHD, Kahler AL, Sunderland, eds.) Pp 43-63.

Hamwiah A, Farah J, Moussally S, Al-Shamaa K, Elmeer K, Khierallah H, Udupa S, Lababidi S, Malek JA, Aaouine M, Baum M (2010). Development of 1000 microsatellite markers across the date palm (*Phoenix dactylifera* L.) genome. *Acta Hort* 882: 269-277.

Hancock JF (2005). Contributions of domesticated plant studies to our understanding of plant evolution. *Ann. Bot.* 96: 953-963.

Harley MM & Baker WJ, (2001). Pollen aperture morphology in Arecaceae: Application within phylogenetic analyses and a summary of fossil records for palm-like pollen. *Grana* 40: 45-77.

- Harley MM (2006). A summary of fossil records for Arecaceae. *Botanical Journal of the Linnean Society* 151: 39-67.
- Henderson S, Billotte N, & Pintaud JC (2006). Genetic isolation of Cape Verde Islands *Phoenix atlantica* (Arecaceae) revealed by microsatellite markers. *Conservation Genetics* 7: 213-223.
- Heuertz M, De Paoli E, Kallman T, Larsson H, Jurman I, Morgante M, Lascoux M, Gyllenstrand N. (2006). Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics* 174: 2095-2105.
- Hewitt GM (1999). Post-glacial re-colonization of European biota. *Biological J Linn Soc* 68: 87-112.
- Horn JW, Fisher JB, Tomlinson PB, Lewis CE, Laubengayer K (2009). Evolution of lamina anatomy in the palm family (Arecaceae). *American Journal of Botany* 96: 1462-1486.
- Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH (2003). From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Can J Bot* 81: 1247-1266.
- Ingvarsson PK, Garcia MV, Hall D, Luquez V, Jansson S (2006). Clinal variation in phyB2, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (*Populus tremula*). *Genetics* 172: 1845-1853.
- Ingvarsson PK (2008). Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics* 180: 329-340.
- Innan H & Kim Y (2004). Pattern of polymorphism after strong artificial selection in a domestication event. *Proc. Natl. Acad. Sci. U. S. A.* 101: 10667-10672.
- IPCC (2007a). *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, UK: Cambridge University Press.
- IPCC (2007b). *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the*

Intergovernmental Panel on Climate Change. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Isabel N, Tremblay L, Michaud M, Tremblay FM, Bousquet J (1993) RAPDs as an aid to evaluate the genetic integrity of somatic embryogenesis-derived populations of *Picea mariana* (Mill.) BSP. *Theor Appl Gene* 86: 81-87.

Jakob SS, Ihlow A, Blattner FR (2007). Combined ecological niche modelling and molecular phylogeography revealed the evolutionary history of *Hordeum marinum* (Poaceae) niche differentiation, loss of genetic diversity, and speciation in Mediterranean Quaternary refugia. *Mol Ecol* 16: 1713-1727.

Jakobsson M & Rosenberg NA (2007). CLUMPP: a Cluster Matching and Permutation Program for Dealing with Label Switching and Multimodality in Analysis of Population Structure. *Bioinformatics* 23: 1801-1806.

Jaramillo-Correa JP, Grivet D, Terrab A, Kurt Y, De-Lucas AI., Wahid N, Vendramin GG, González-Martínez SC (2010). The Strait of Gibraltar as a major biogeographic barrier in Mediterranean conifers: a comparative phylogeographic survey. *Molecular Ecology* 19: 5452-5468.

Jarne P & Lagoda PJJ (1996). Microsatellites, from molecules to populations and back. *Trends Ecol Evol* 11: 424-429.

Johnson C, Cullis TA, Cullis MA, Cullis CA (2009). DNA markers for variety identification in date palm (*Phoenix dactylifera* L.). *Journal Horticult Sci Biotech* 84: 591-594.

Kalinowski ST, Taper ML, Marshall TC (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol* 16: 1099-1006.

Kärkkäinen K & Savolainen O (1993). The degree of early inbreeding depression determines the selfing rate at the seed stage: model and results from *Pinus sylvestris* (Scots pine). *Heredity* 71: 160-166.

Kashi Y, King D, Soller M (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet* 13: 74-78.

Kaufmann MR (1996). To live fast or not: growth, vigor and longevity of old-growth ponderosa pine and logpole pine trees. *Tree Physiol* 16: 139-144.

Keely JE & Zedler PH (1998). Evolution of life histories in *Pinus*. In: Richardson DM (ed) Ecology and Biogeography of *Pinus*, Cambridge University Press, Cambridge. pp. 219-251.

Kinlaw CS & Neale DB (1997). Complex gene families in pine genomes. Trends Plant Sci 2(9): 356-359.

Klaus W (1989). Mediterranean pines and their history. Plant Syst Evol 162: 133-163.

Koch MA, Haubold B, Mitchell-Olds T (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol Biol Evol 17: 1483-1498.

Koelewijn HP, Koski V, Savolainen O (1999). Magnitude and timing of inbreeding depression in Scots pine (*Pinus sylvestris* L.). Evol 53: 758-768.

Koski V (1970). A study of pollen dispersal as a mechanism of gene flow in conifers. Communicationes Instituti Forestalis Fenniae 70. pg 78.

Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein EK, Ritland K, Kuparinen A, Gerber S, Schueler S (2012). Long-distance gene flow and adaptation of forest trees to rapid climate change. Ecology Letters 15: 378-392.

Krupkin AB, Liston A, Strauss SH (1996). Phylogenetic analysis of the hard pines (*Pinus* subgenus *Pinus*, Pinaceae) from chloroplast DNA restriction site analysis. Am J Bot 83: 489-498.

Krutovskii KV & Neale DB (2001). Forest genomics for conserving adaptive genetic diversity. Forest genetic resource working paper. Forestry Department, FAO, Rome. Pp. 1-30.

Krutovsky KV, Troggio VM, Brown GR, Jermstad KD, Neale DB (2004). Comparative mapping in the Pinaceae. Genetics 168: 447-461.

Kusumi J, Tsumura Y, Yoshimaru H, Tachida H (2002). Molecular evolution of nuclear genes in *Cupressaceae*, a group of conifer trees. Mol Biol Evol 19: 736-747.

Lai Y & Sun F (2003). The relationship between microsatellites slippage mutation rate and the number of repeat units. Mol Biol Evol 20: 2123-2131.

- Langella O (2002). Populations 1. 2. 28 Software. CNRS. UPR9034. France.
- Larkin PJ & Scowcroft WR (1981). Somatoclonal variations – a novel source of variability from cell cultures for plant improvement. *Theor Appl Genetics* 60: 197-214.
- Le Corre V & Kremer A (2003). Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics* 164: 1205–1219.
- Levinson G & Gutman A (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4: 203-221.
- Librado P & Rozas J (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.
- Liber Z, Nikolic T, Mitic B, Satovic Z (2003). RAPD markers and black pine (*Pinus nigra* arnold) intraspecies taxonomy – evidence from the study of nine populations. *Acta Societatis Botanicorum Poloniae* 72: 249-257.
- Liston A, Robinson WA, Pinero D, Alvarez-Buylla ER (1999). Phylogenetics of *Pinus* (Pinaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. *Mol Phylogenet Evol* 11: 95-109.
- Little EL & Critchfield WB (1969). Subdivisions of the genus *Pinus* (pines). USDA Forest Service, Washington DC, Miscellaneous Publication Number 1144.
- Lucic A, Mladenovic-Drinic S, Stavretovic N, Isajev V, Lavadinovic V, Rakonjac L, Novakovic M (2010). Genetic diversity of Austrian pine (*Pinus nigra* Arnold) populations in Serbia revealed by RAPD. *Archives of Biological Sciences* 62: 329-336.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature* 4: 981-994.
- Lumaret R, Ouazzani N, Michaud H, Vivier G, Deguilloux MF, Di Giusto F (2004). Allozyme variation of oleaster populations (wild olive tree) (*Olea europea* L.) in the Mediterranean Basin. *Heredity* 92: 343-351.

- Ma XF, Szmidt AE, Wang XR (2006). Genetic structure and evolutionary history of a diploid hybrid pine *Pinus densata* inferred from the nucleotide variation at seven gene loci. *Mol Bio Evol* 23(4): 807-816.
- MacDonald GM (1993). Fossil pollen analysis and the reconstruction of plant invasions. *Adv Ecol Res* 24: 67–110.
- Mai DH (1989). Systematics and evolution development and regional differentiation of the European vegetation during the Tertiary. *Plant System Evol* 162: 79-91.
- Malek JA (2010). Next generation DNA sequencing applied to the date palm tree (*Phoenix dactylifera* L.). *Acta Horticulture* 882: 249-252.
- Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* 7: 639-655.
- McKey D., Elias M., Pujol B, Duputié A (2010). The evolutionary ecology of clonally propagated domesticated plants. *New Phytologist* 186 : 318-332.
- McLachlan JS, Clark JS, Manos PS (2005). Molecular indicator of tree migration capacity under rapid climate change. *Ecology* 86(8): 2088-2098.
- McNally KL et al. (2009). Genomewide SNP variation reveals relationships among landrace and modern variety of rice. *Proc Natl Acad Sci USA* 106: 12273-12278.
- Menzel A & Fabian (1999). Growing season extended in Europe. *Nature* 397: 659-659.
- Meshinev T, Apostolova I, Koleva E (2000). Influence of warming on timberline rising: a case study on *Pinus peuce* Griseb. in Bulgaria. *Phytocoenologia* 30: 431-38.
- Menzel A, Sparks TH, Estrella N, Koch E, Aasa A, Ahas R, Alm-Kubler K, Bissolli P, Braslavská O, Briede A, Chmielewski FM et al. (2006). European phenological response to climate change matches the warming pattern. *Global Change Biology* 12:1969-1976.
- Millar CI (1993). Impact of the Eocene on the evolution of *Pinus* L. *Ann. Missouri Bot. Gard* 80: 471-498.

Millar CI (1998). Early evolution of pines. In: Richardson DM (ed) *Ecology and Biogeography of Pinus*. Cambridge Univ Press, Cambridge, UK, pp 69-91.

Miller AJ & Gross BL (2011). From forest to field: Perennial Fruit Crop Domestication. *Am J Botany* 98: 1389-1414.

Mitchell SE, Kresovich S, Jester CA, et al. (1997). Application of multiplex PCR and fluorescence- based, semi-automated allele sizing technology for genotyping plant genetics resources. *Crop Science* 37: 617-624.

Moore HE (1963). An annotated checklist of cultivated palms. *Principes* 7: 119-182.

Moore HE & Uhl NW (1982). Major trends of evolution in palms. *Botanical review* 48: 1-69.

Moraga AR, Perez DC, Borja MEL, Tiscar PA, Vinegla B, Linares JC, Gomez LG, Ahrazem O (2012). Genetic diversity of *Pinus nigra* Arn population in Southern Spain and Northern Morocco revealed by Inter sequence repeat profiles. *Int J Mol Sci* 13: 5645-5658.

Morgante M, Hanafey M, Powell W (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature* 30: 194-200.

Morgante M & Olivieri A (1993). PCR-amplified microsatellites as markers in plant genetics. *Plant Journal* 3: 175-182.

Morgante M & Vendramin GG (1991). Genetic variation in Italian population of *Picea abies* L. Karst and *Pinus leucodermis* Ant. In: *Genetic Variation in European Population of Forest trees*. (Muller-Starck G & Ziehe M, eds.). Frankfurt, pp. 205-227.

Morgenstern EK (1996). *Geographic variation in forest trees: genetic basis and application of knowledge in silviculture*. UBC Press, Vancouver, Canada. Pg 206.

Mosca E, Eckert AJ, Liechty JD, Wegrzyn JL, La Porta N, Vendramin GG, Neale DB (2012). Contrasting patterns of nucleotide diversity for four dominant conifers of montane European forests. *Evol Appl* 5(7): 762-775.

Muona O & Harju A (1989). Effective population sizes, genetic variability and mating system in natural stands and seed orchards of *Pinus sylvestris* Sil Gen 38: 221-228.

Munier P (1974). Le problème de l'origine du palmier-dattier et l'Atlantide. Fruits 29: 235-240.

Munier P (1973). Le Palmier Dattier-Techniques agricoles et productions tropicales. Maison neuve & Larose. Paris.

Neale DB & Ingvarsson PK (2008). Population, quantitative and comparative genomics of adaptation in forest trees. Curr Opin Plant Biol 11: 149-155.

Neale DB & Kremer A (2011). Forest tree genomics: growing resource and applications. Nature 12: 111-122.

Neale DB & Savolainen O (2004). Association genetics of complex traits in conifers. Trends Plant Sci 9: 325-330.

Neale DB & Sederoff RR (1989). Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. Theor Appl Genet 77: 212-216.

Nei M, Tajima F, and Tatenno Y, 1983. Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19:153-170.

Nei M (1987). Molecular evolutionary genetics. Columbia Univ. Press.

Nikolić D & Tucić N (1983). Isoenzyme variation within and among populations of European black pine (*Pinus nigra* Arnold). Silv Gen. 32: 80-89.

Nixon RW (1936). Metaxenia and interspecific pollinations in *Phoenix*. American Society for Horticultural Science 33: 21-26.

Nixon RW & Furr JR (1965). Problem and progress in date breeding. Date Growers' Inst. Rep. 42: 2-5.

Nordborg M & Weigel D (2008) Next-generation genetics in plants. Nature 456(7223): 720-723.

Palme AE, Wright M, Savolainen O (2008). Patterns of divergence among conifer ESTs and polymorphism in *Pinus sylvestris* identify putative selective sweeps. *Mol Biol Evol* 25: 2567-2577.

Palme AE, Pyhajarvi T, Wachowiak W, Savolainen O (2009). Selection on nuclear genes in a *Pinus* phylogeny. *Mol Biol Evol* 26:893-905.

Parks DH, Porter M, Churcher S, Wang S, Blouin C, Whalley J, Brooks S, Beiko RG (2009). GenGIS: A geospatial information system for genomic data. *Genome Research* 19: 1896-1904.

Peakall R, Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6: 288-295.

Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Muller-Starck G, Demesure-Musch B, Palme A, Martin JP, Rendell S, Vendramin GG (2003). Glacial refugia: hotspot but not melting pots of genetic diversity. *Science* 300: 1563-1565.

Petit RJ, Hampe A, Cheddadi R (2005). Climate change and tree phylogeography in the Mediterranean. *Taxon* 54: 877-885.

Pintaud JC, Zehdi S, Couvreur T, Barrow S, Henderson S, Bertossi FA, Tregear J, Billotte N (2010). Species Delimitation in the Genus Phoenix (Arecaceae) Based on SSR Markers, with Emphasis on the Identity of the Date Palm (*Phoenix dactylifera*). In Seberg, Petersen, Barford & Davis, eds. *Diversity, Phylogeny, and Evolution in the Monocotyledons*. Aarhus University Press, Denmark, 2010. Pp.267-286.

Pitelka LF, Gardner RH, Ash J, Berry S, Gitay H, Noble IR, Saunders A, Bradshaw RHW, Brubaker L, Clark JS, Davis MB, Sugita S, Dyer JM, Hengeveld R, Hope G, Huntley B, King GA, Lavorel S, Mack RN, Malanson GP, McGlone M, Prentice IC, Rejmanek M (1997) Plant migration and climate change. *Am Sci* 85: 464–473.

Plomion C, Chagné D, Pot D, Kumar S, Wilcox PL, Burdon RD, Prat D, Peterson DG, Paiva J, Chaumeil P, Vendramin GG, Sebastiani F, Nelson CD, Echt CS, Savolainen O, Kubisiak TL, Cervera MT, de María N, Islam-Faridi MN (2006). The Pines. In: *Genome Mapping and Molecular Breeding in Plants*, Vol. 7 Forest Trees, Chitta R. Kole (Ed). Springer, Heidelberg, Berlin, New York, Tokyo. Pg 82.

Popenoe P (1973). *The date palm*. Edited by Henry Field. Miami: Field Research Projects.

Powell W, Morgante M, Andre C, McNicol JW, Machray GC, Doyle JJ, Tingey SV, Rafalski JA (1995). Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Curr Biol* 5: 1023-1029.

Price RA, Liston A, Strauss SH (1998). Phylogeny and systematics of *Pinus*. In: Richardson DM (ed) *Ecology and Biogeography of Pinus*. Cambridge Univ Press, Cambridge, UK, pp 49-68.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Provan J, Powell W, Hollingsworth PM (2001). Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol Evol* 16:142-147.

Provan J & Bennett KV (2008). Phylogeographic insight into cryptic glacial refugia. *Trends Ecol Evol* 23(10): 564-571.

Pyhäjärvi T, Garcia-Gil MR, Knutt T, Mikkonen M, Wachowiak W, Savolainen O (2007). Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics* 177: 1713-1724.

Pot D, McMillan L, Echt C, Le Provost G, Garnier-Géré P, Cato S, Plomion C (2005). Nucleotide variation in genes involved in wood formation in two pine species. *New Phytol* 167: 101-112.

Quézel P. & Médail F (2003). *Ecologie et Biogéographie des Forêts du Bassin Méditerranéen*. Elsevier, Paris.

Rafii ZA, Dodd R, Zavarin E (1996). Genetic diversity in foliar terpenoids among natural populations of European black pine. *Bioch Syst Ecol* 24: 325-339.

Ranade K, Chang MS, Ting CT, Pei D, Hsiao CF, et al., (2001). High-throughput genotyping with single nucleotide polymorphisms. *Genome Res*. 11: 1262-1268.

Ream CL (1976). Metaxenia effect of pollen from inbreed male palms on ripening period and size of date fruit. *Date Grovers' Inst. Rep.* 53: 21-22.

Renner SS & Ricklefs RE (1995). Dioecy and its correlates in the flowering plants. *Am J Botany* 82: 596-606.

Rehfeldt GE, Tchebakova NM, Parfenova YI, Wykoff WR, Kuzmina NA, Milyutin LI (2002). Intraspecific responses to climate in *Pinus sylvestris*. *Global Change Biol* 8: 912-929.

Reille M, Gamisans J, Beaulieu JL, Andrieu V (1997). The late-glacial at Lac de Creno (Corsica, France): a key site in the western Mediterranean basin. *New Phytologist* 135: 547-559.

Rhouma A (1999). A propos de la culture du palmier-dattier en Tunisie. Colloque sur la production, la confection et la distribution des dattes entre la Tunisie et le Sultanat d'Oman. Université du Sultan Qabous, Faculté d'Agronomie, Oman.

Rhouma S, Zehdi SA, Salem AOM, Rhouma A, Marrakchi M, Trifi M (2007). Genetic diversity in ecotypes of Tunisian date palm (*Phoenix dactylifera* L.) assessed by AFLP markers. *J Hort Sci Biotech* 82: 929-933.

Rhouma S, Baraket G, Dakhlaoui DS, Zehdi SA, Trifi M (2011). Molecular research on the genetic diversity of Tunisian date palm (*Phoenix dactylifera* L.) using the random amplified microsatellite polymorphism (RAMPO) and amplified fragment length polymorphism (AFLP) methods. *African Journal of Biotechnology* 10: 10352-10365.

Richardson DM & Rundel PW (1998). Ecology and biogeography of *Pinus*: an introduction. In: Richardson DM (ed) *Ecology and Biogeography of Pinus*. Cambridge Univ Press, Cambridge, UK, pp 3-46

Rival A, Tregear J, Verdeil JL, Richaud F, Beule T, Duval Y, Hartmann C (1998). Molecular search for mRNA and genomic marker of the oil palm "mantled" somaclonal variation. *Acta Hort.* 461: 165-171.

Rokas AB, Williams L, King N, Carroll S B (2003). Genome scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425: 798-804.

Roseberg NA, Pritchard JK, Weber JL (2002). Genetic structure of human populations. *Science* 298: 2381-2385.

Saaidi M, Toutain G, Bannerot H, Louvet J (1981). La sélection du palmier dattier (*Phoenix dactylifera* L.) pour la résistance au Bayoud. *Fruits* 36: 241-249.

Saaidi M (1992). Comportement au champ de 32 cultivars vis-a-vis du Bayoud. 25 années d'observations. *Agronomie* 12: 369-378.

Saker MM, Bekheet SA, Taha HS, Fahmy AS, Moursy HA (2000). Detection of somaclonal variation in tissue culture-derived date palm plants using isoenzyme analysis and RAPD fingerprints. *Biologia Plantarum* 43: 347-351.

Salem AOM, Rouma S, Zehdi S, Marrakchi M, Trifi M (2008). Morphological variability of Mauritanian date palm (*Phoenix dactylifera* L.) cultivars revealed by vegetative traits. *Acat Bot. Croat.* 67: 81-90.

Scarin E (1937). La Giofra e Zella (Le oasi del 29° parallelo della libia occidentale). *Rivista geografica italiana* V-VI. Firenze, Italia.

Sannikov SN & Petrova IV (2012). Phylogenogeography and Genotaxonomy of *Pinus sylvestris* L. populations. *Russ J Ecol* 43(4): 273-280.

Savolainen O & Pyhäjärvi T (2007). Genomic diversity in forest trees. *Curr Opin Plant Biol* 10: 162-167.

Savolainen O, Pyhäjärvi T, Knurr T (2007). Gene flow and local adaptation in trees. *Ann. Rev. Ecol. Evol. Syst.*, 38: 595-619.

Scaltsoyiannes A, Tsaktsira M, Pasagiannis G, Tsoulpha P, Zhelev P, Iliev I, Rohr R (2009). Allozyme variation of European Black (*Pinus nigra* Arnold) and Scots pine (*Pinus sylvestris* L.) populations and implications on their evolution: a comparative study. *Journal of Biological Research - Thessaloniki* 11: 95-106.

Schlotterer C (2004). The evolution of molecular markers-just a matter of fashion? *Nature Genetics* 5: 63-69.

Schuelke M (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*. 18: 233-234.

Schulman E (1958). Bristlecone pine, oldest known living thing. *Natl Geogr Mag.* 113: 354-372.

Schuster WSF & Mitton JB (2000). Paternity and gene dispersal in limber pine (*Pinus flexilis* James). *Heredity* 84: 348–361.

Schütt P (1959). Züchtung mit Kiefern Teil 2. Kreuzungen. Resistenzzüchtung und zytologie. Mitteilungen der Bundesforschungsanstalt fuer Forst-und Holzwirtschaft 42. Forstgenetik und Forstpflanzenzüchtung. p. 1-40.

Sedra MH, Lashermes P, Trouslot P, Combes M, Hamon S (1998). Identification and genetic diversity analysis of date palm (*Phoenix dactylifera* L.) varieties from Morocco using RAPD markers. *Euphytica* 103:75–82

Seifert S, Vornam B, Finkeldey R (2012). DNA sequence variation and development of SNP markers in beech (*Fagus sylvatica* L.). *European Journal of Forest Research* 131: 1761-1770.

Shiraishi S & Shiraishi S (2011). Nucleotide diversities and genetic relationship in the three Japanese Pine Species; *Pinus thunbergii*, *Pinus densiflora*, and *Pinus luchuensis*. *Diversity* 3(1):121-135.

Siljak-Yakovlev S, Benmalek S, Cerbah M, Coba de la Pena T, Bounaga N, Brown SC, Sarr A (1996). Chromosomal sex determination and heterochromatin structure in date palm. *Sex Plant Reproduction* 9: 127-132.

Siol M, Wright SI, Barrett CH (2010). The population genomics of plant adaptation. *New Phytologist* 188: 313-332.

Small RL, Cronn RC, Wendel JF (2004). Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17: 145-170.

Soliman SS, Ali BA, Ahmed MMM (2003). Genetic comparisons of Egyptian date palm cultivars (*Phoenix dactylifera* L.) by RAPDPCR. *Afr J Biotechnol* 2: 86-87.

Soniya EV, Banerjee NS, Das MR (2001). Genetic analysis of somaclonal variation among callus-derived plants of tomato. *Cur Sci* 80: 1213-1215.

Sorensen FC (2001). Effect of population outcrossing rate on inbreeding depression in *Pinus contorta* var. *murrayana* seedlings. *Scandinavian Journal of Forest Research* 16: 391-403.

Soto A, Robledo-Arnuncio JJ, González-Martínez SC, Smouse PE, Alía R (2010). Climatic niche and neutral genetic diversity of the six Iberian pine species: a retrospective and prospective view. *Mol Ecol* 19(7): 1396-1409.

Stockey RA, Kvacek J, Hill R, Rothwell GH, Kvacek Z (2005). The fossil record of the Cupressaceae. In A Farjon, ed, *A Monograph of the Cupressaceae and Sciadopitys*. Royal Botanic Gardens, Kew, UK, pp -68.

Strand M, Prolla TA, Liskay RM, Petes TD (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 365: 274-276.

Strauss SH & Doerksen AH (1990). Restriction fragment analysis of pine phylogeny. *Evolution* 44: 1081-1096.

Stewart WN & Rothwell GW (1993). *Paleobotany and the evolution of plants*, 2nd ed. Cambridge University Press, Cambridge, UK.

Stephens M, Smith N, Donnelly P (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* : 978-989.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998). Comparative phylogeography and post glacial colonization routes in Europe. *Mol Ecol* 7: 453-464.

Takrouni L, Rhouma A, Khoualdia O, Allouchi B (1988). Observations préliminaires sur deux graves "maladies" d'origine inconnue du palmier dattier en Tunisie. *Ann. Inst. Nat. Rech. Agr. de Tunisie* 61.

Tajima F (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.

Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Tamura K (1999). *TreeExplorer* Version 2.12. Pennsylvania State University, University Park, PA.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: Molecular evolutionary genetics analysis using likelihood, distance, and parsimony methods. *Mol Biol Evol* 28(10): 2731-2739

Tautz D & Schlotterer C (1994). Concerted evolution, Molecular Drive and Natural-Selection-Reply. *Curr Biol* 4: 1166-1166.

Temesgen B, Brown GR, Harry DE, Kinlaw CS, Sewell MM, Neale DB (2001). Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda*). *Theor Appl Gen* 102: 664-675.

Terral JF, Alonso F, Capdevilla RBI, Chatti N, Fabre L, Fiorentino G, Marinval P, et al. (2004). Historical biogeography of olive domestication (*Olea europaea* L.) as revealed by geometrical morphometry applied to biological and archeological material. *Journal of Biogeography* 31: 63-77.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* 24: 4876–4882.

Tisserat B (1979). Propagation of date palm (*Phoenix dactylifera*) in vitro. *Journal Exp. Bot.* 30: 1275-1283.

Tolun A, Velioglu E, Cengel B, Kaya Z (2000). Genetic structure of Black Pine (*Pinus nigra* Arnold subspecies *pallasiana*) populations sampled from the Bolkar Mountains. *Silv Gen* 49: 113-119.

Torres AM & Tisserat B (1980). Leaf isozymes as genetic markers in date palm. *Am Jour Bot* 67: 162-167.

Travis JMJ, Munkemuller T, Burton OJ (2010). Mutation surfing and the evolution of dispersal during range expansions. *J Evol Biol* 12: 2656–2667.

Trifi M, Rhouma A, Marrakchi M (2000) Phylogenetic relationships in Tunisian date-palm (*Phoenix dactylifera* L.) germplasm collection using DNA amplification fingerprinting. *Agronomie* 20: 665-671.

Triki MA, Zouba A, Khouldia O, Ben Mahamoud O, Takrouni ML, Garnier M, Bové JM, Montarone M, Poupet A, Flores R, Darós JA, Fadda ZGN, Moreno P, Duran-Vila N, (2003). “Maladie des Feuilles Cassantes” or Brittle Leaf Disease of date palms in Tunisia: biotic or abiotic disease? *Journal of Plant Pathology* 85: 71-79.

Van der Burgh J (1973). Hölzer der niederrheinischen Braunkohlenformation. 2. Hölzer der Braunkohlengruben 'Maria Theresia' zu Herzogenrath, 'Zukunft West' zu Eschweiler und 'Victor' (Zülpich Mitte)

zu Zülpich. Nebst einer systematisch-anatomischen Bearbeitung der Gattung Pinus L. Rev Palaeobot Palynol 15: 73-275.

Vendramin GG, Lelli L, Rossi P, Morgante M (1996). A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. Mol Ecol 5: 595-598.

Vendramin GG, Fady B, González-Martínez SC et. al. (2008) Genetically depauperate but widespread: the case of an emblematic Mediterranean pine. Evolution 62: 680–688.

Vidakovic M (1958). Investigations on the intermediate type between the Austrian and Scots pine. Silv Gen 7:12-19.

Wachowiak W, Balk PA, Savolainen O (2009). Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (*Pinus sylvestris* L.). Tree Gen Genomes 5: 117-132.

Wachowiak W, Palme AE, Savolainen O (2011). Speciation history of three closely related pines *Pinus mugo* (T.), *P. uliginosa* (N.) and *P. sylvestris* (L.). Mol Ecol. 20:1729-1743.

Wang XR, Szmidt AE, Lindgren D (1991). Allozyme differentiation among populations of *Pinus sylvestris* (L.) from Sweden and China. Hereditas 114: 219-226.

Wang XT, Tsumara Y, Yoshimaru H, Nagasaka K, Szmidt A (1999). Phylogenetic relationships of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *matK*, *rpl20-rps18* spacer, and *trnV* intron sequences. Am J Bot 86: 1742-1753.

Watterson GA (1975). On the number of segregating sites in genetical models without recombination. Theor. Pop. Biol. 7: 256-276.

Weising K & Gardner RC (1999). A set of conserved PCR primers for the analysis of simple sequence repeat polymorphism in the chloroplast genomes of dicotyledonous plant. Genome 42: 9-19.

Willyard A, Syring J, Gernandt DS, Liston A, Cronn R (2007). Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. Mol Biol Evol 24: 90-101.

- Wilson MA, Gaut B & Clegg MT (1990). Chloroplast DNA evolves slowly in the palm family (Arecaceae). *Mol Biol Evol* 7: 303-314.
- Wrigley G (1995). Date-palm (*Phoenix dactylifera* L.). In: The evolution of crop plants. (Eds.): J. Smartt and N.W. Simmonds. London: Longman, pp. 399-403.
- Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, Zhao D, Al-Mssallem IS, Yu J (2010). The Complete Chloroplast Genome Sequence of Date Palm (*Phoenix dactylifera* L.). *PLoS ONE* 5(9): e12762.
- YinY, Zhang X, Fang Y et al (2012) High-throughput sequencing based gene profiling on multi-staged fruit development of date palm (*Phoenix dactylifera* L.). *Plant Mol Biol* 78(6): 617–626.
- Zaid A & de Wet PF (2002a). Botanical and systematic description of the date palm. In: Zaid A (ed). Date palm cultivation. FAO Plant Production and Protection Paper no. 156. Food and Agriculture Organisation of the United Nations, Rome, pp 30–37.
- Zaid A & de Wet PF (2002b). Origin, geographical distribution and nutritional values of date palm. In: Zaid A (ed). Date palm cultivation. FAO Plant Production and Protection Paper no. 156. Food and Agriculture Organisation of the United Nations, Rome, pp 7–29.
- Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Mol Ecol* 12: 563-584.
- Zehdi S, Sakka H, Rhouma A, Salem AOM, Marrakchi M, Trifi M (2004a). Analysis of Tunisian date palmgermplasm using simple sequence repeat primers. *Afri J Biotech* 3: 215-219.
- Zehdi S, Trifi M, Billotte N, Marrakchi M, Pintaud JC (2004b). Genetic diversity of Tunisian date palms (*Phoenix dactylifera* L.) revealed by nuclear microsatellite polymorphism. *Hereditas* 141: 278-287
- Zohary D & Hopf M (2000). Domestication of plants in the Old World. The origin and spread of cultivated plants in West Asia, Europe and the Nile Valley. 3rd Ed. New York: Oxford University Press.
- Zohary D & Spiegel-Roy P (1975). Beginnings of fruit growing in the Old World. *Science* 187: 319-327.