



Università di Pisa



Dipartimento di Scienze Agrarie, Alimentari e Agro-Ambientali

Corso di Laurea Magistrale

in

Biotecnologie Vegetali e Microbiche

Caratterizzazione geografica, molecolare e
fenotipica di una collezione di popolazioni naturali
Triticum urartu, progenitore del frumento
moderno

Relatori

Dott. Rodolfo Bernardi

Prof. Mario Enrico Pè

Candidato

Rebecca Fiorella Talini

Anno Accademico

2014/2015

Indice

Riassunto	4
1 Introduzione	6
1.1 La storia evolutiva e l'importanza del frumento nel mondo	6
1.2 Sfide dell'agricoltura moderna.....	14
1.3 Il progenitore selvatico <i>Triticum urartu</i>	24
1.4 Strategie di indagine della diversità di <i>Triticum urartu</i>	27
2 Scopo della tesi	32
3 Materiale e metodi	34
3.1 La collezione	34
3.2 Analisi GIS.....	35
3.2.1 Gestione dati GIS.....	37
3.3 Germinazione del materiale ed estrazione del DNA.....	39
3.3.1 Camera di crescita.....	39
3.3.2 Estrazione del DNA	39
3.3.3 Controllo qualità e quantità del DNA.....	42
3.3.4 Normalizzazione dei campioni.....	43
3.4 Genotipizzazione RADSeq	44
3.4.1 Analisi dati molecolari.....	48
3.5 Campo sperimentale.....	49
3.5.1 La semina.....	49
3.5.2 Analisi fenotipiche	50
3.5.3 Gestione dati fenotipici.....	50
4 Risultati	52
4.1 Geografici.....	52
4.2 Risultati molecolari	65
4.3 Risultati fenotipici.....	71
4.3.1 Il campo sperimentale.....	71
4.3.2 Risultati dati fenotipici.....	76
5 Discussione	85
5.1 Distribuzione ecologica di <i>Triticum urartu</i>	85
5.2 Diversità molecolare in <i>Triticum urartu</i>	89

5.3	Caratterizzazione fenotipica.....	92
6	Conclusioni.....	97
7	Ringraziamenti	98
8	Bibliografia.....	99

Riassunto

Il lavoro di tesi è stato incentrato sulla caratterizzazione a diversi livelli di una collezione di *Triticum urartu*, progenitore selvatico e donatore del genoma A del frumento moderno. Questo studio può essere considerato come l'inizio di un percorso volto a scoprire loci genomici coinvolti nell'adattamento ambientale attraverso la *landscape genomics*, una tecnica che fonde la genomica di popolazione e l'ecologia.

Triticum urartu può essere considerato una fonte inesplorata di nuova variabilità allelica per il miglioramento di frumento. Essendo selvatico, *T. urartu* non ha subito alcun processo di domesticazione e pertanto non è andato incontro alla forte riduzione di variabilità allelica che hanno subito le specie di frumento coltivate. Il genoma di *T. urartu* è risultato della sola selezione naturale, e possiede i caratteri di rusticità che gli conferiscono la resistenza e resilienza a vari tipi di *stress* (siccità, alte temperature, malattie, ecc.), e che rendono possibile il suo adattamento alle differenti condizioni ambientali a cui è sottoposto. La scelta di questo progenitore selvatico è anche motivata dal recente sequenziamento del suo genoma, che renderà più efficiente il trasferimento di questi caratteri al frumento tramite l'omeologia dei genomi di queste due specie.

Con l'obiettivo di osservare la più elevata variabilità genetica possibile, sono state aggregate 428 accessioni che rappresentano la reale distribuzione di *T. urartu* nella Mezzaluna Fertile. Il lavoro di tesi si è quindi articolato su 3 livelli di caratterizzazione: i) geografica, studiando le variabili ambientali relative ai siti di campionamento dell'intera collezione, e quindi valutando le condizioni climatiche a cui ogni accessione è sottoposta; ii) molecolare, utilizzando una tecnica di *next-generation sequencing* su un

subset di 24 individui per valutare la diversità della collezione; iii) fenotipica, valutando la distribuzione di diversi caratteri nella collezione di *T. urartu* cresciuta in campo.

I risultati ottenuti dall'analisi geografica hanno evidenziato l'elevata variabilità ambientale ed ecologica presente nei siti di campionamento della collezione. Le accessioni sono distribuite attraverso un range molto ampio di condizioni climatiche, in aree che presentano altitudini, temperature e piovosità altamente differenti.

L'analisi della diversità molecolare ha rilevato all'interno della collezione un'elevata variabilità, che non risulta essere influenzata dalla distanza geografica. Inoltre, la varianza molecolare del *daset* è scarsamente strutturata e difficilmente condensabile in un numero limitato di variabili.

I rilievi in campo hanno mostrato l'altrettanto elevata diversità fenotipica raccolta nella collezione. I principali tratti agronomici misurati (altezza della pianta, area foglia bandiera, lunghezza spiga, ecc.) presentano una distribuzione normale, dimostrando le potenzialità di questa risorsa in un'ottica di genetica quantitativa.

1 Introduzione

1.1 La storia evolutiva e l'importanza del frumento nel mondo

Il frumento è la coltura più diffusa nel mondo (Feldman *et al.*, 1995; Gustafson *et al.* 2009) dopo il riso e il mais (<http://faostat.fao.org>). Ad oggi, il frumento è coltivato su circa 215 milioni di ettari di terreno, il 16% della superficie agricola globale utilizzata. Annualmente vengono prodotti circa 700 milioni di tonnellate di granella, che provvedono al 20% del fabbisogno calorico umano e delle proteine apportate con la dieta (<http://faostat.fao.org>; Peng, Sun, e Nevo 2011a). Il frumento, da un punto di vista nutritivo, è di gran lunga superiore ad altri cereali (mais, riso, sorgo o orzo), come dimostrato dai valori nutrizionali della sua farina ottenuta dalla macinazione della granella: essa infatti non contiene solamente amido (60-80% di carboidrati), ma presenta anche un buon contenuto di proteine (8-14% di gliadine e glutenine, necessarie a formare il glutine) (Zohary, Hopf, e Weiss 2012).

Il frumento viene spesso percepito come il cereale più rappresentativo dell'agricoltura del vecchio mondo (Zohary e Hopf 2000). E' stato infatti una delle prime piante ad essere coltivate, e la sua domesticazione, avvenuta circa 10,000 anni fa, ha segnato una svolta importante nell'evoluzione e nello sviluppo della civiltà umana (Brown *et al.* 2009). Prima dell'agricoltura gli uomini erano cacciatori e raccoglitori, il loro sostentamento dipendeva unicamente da quello che riuscivano a cacciare o trovare nell'ambiente. Il loro stile di vita nomade, altamente dipendente dalla natura, si basava cioè sulla disponibilità di animali o piante. Con l'avvento dell'agricoltura e dell'allevamento, l'uomo ha potuto esercitare un certo controllo sulle proprie risorse alimentari, proteggendo se stesso dall'incertezza della natura.

È opinione diffusa che la domesticazione del frumento abbia indotto l'uomo ad organizzarsi in società civili sempre più complesse (Diamond 2015). Mentre gli ortaggi potevano essere coltivati intorno a un accampamento costituito da pochi uomini, il frumento, a causa delle sue esigenze colturali, ha spinto i primi agricoltori a creare reti di canali per l'irrigazione, a costruire le prime mura difensive per proteggere il raccolto durante l'anno ed infine ad organizzare eserciti per proteggere le città nate intorno ai campi coltivati (Ambaglio e Foraboschi 1994; Diamond 2015).

Le prime forme di agricoltura sono iniziate, autonomamente, in differenti parti del mondo (Janick e Paul Gepts 2010). Le regioni che hanno avuto un ruolo chiave nella domesticazione delle principali coltivazioni sono sorprendentemente poche, ma di straordinaria importanza: il Mesoamerica, dove è avvenuta la domesticazione dell'erba selvatica teosinte nel mais; la regione dello Yangtze, dove è stato coltivato per la prima volta il riso; le pianure e gli altopiani del Sud America, da dove provengono le patate, le arachidi e la manioca (Ambaglio e Foraboschi 1994). La domesticazione del frumento è invece avvenuta nella Mezzaluna Fertile, una regione che si estende dalle zone montuose dell'Asia minore (odierna Turchia) fino al Golfo Persico, e che comprende anche le valli adiacenti ai fiumi Tigri, Eufrate e Giordano (Liverani 2009) (Figura 1).

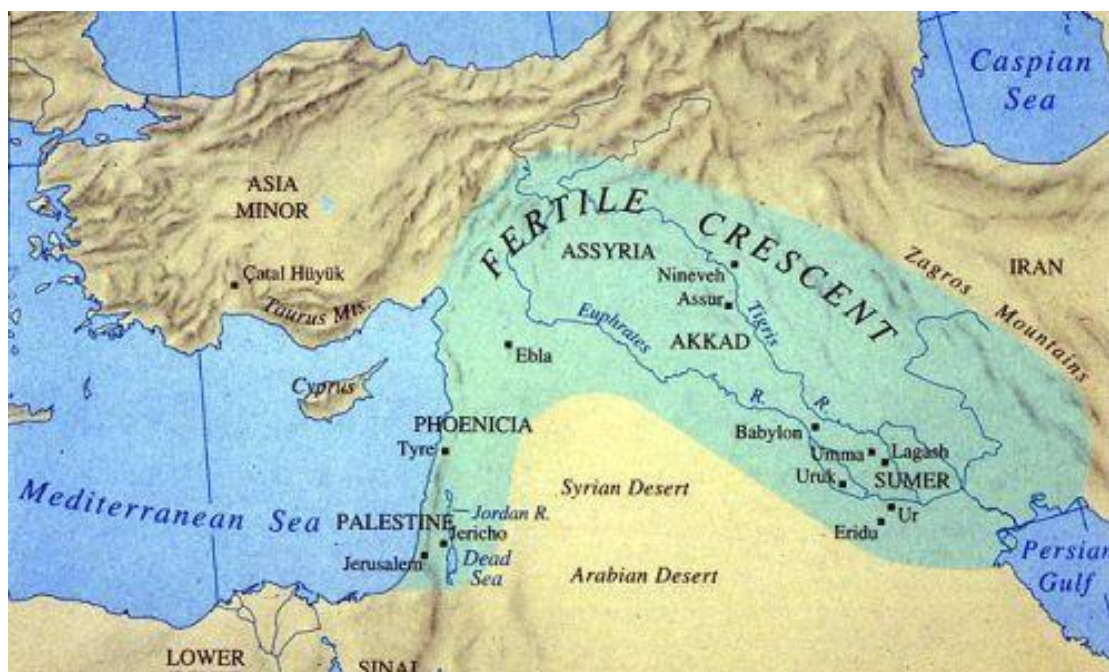


Figura 1. Cartina che rappresenta la ‘Mezzaluna Fertile’, luogo dove è avvenuta la domesticazione del frumento. La regione si estende dalle zone montuose dell’Asia Minore fino al Golfo Persico, e comprende le valli adiacenti ai fiumi Tigri, Eufrate e Giordano, (<http://www.greenme.it/>).

La presenza di questi grandi fiumi, che rendeva fertili le terre, diede l’*input* iniziale alla costruzione di dighe, canali e città; gettando in questo modo le basi dell’agricoltura (Figura 2), della domesticazione del frumento e della sua evoluzione fino ai nostri giorni (Ambaglio e Foraboschi 1994). Oltre alla fertilità delle terre, ciò che ha reso la Mezzaluna Fertile la culla del frumento è stata la presenza delle specie progenitrici di quelle coltivate, come il frumento selvatico, che erano già abbondanti e produttive in natura (Diamond 2015).



Figura 2. La raccolta del grano nell'antico Egitto in una raffigurazione murale, (www.dailymail.co.uk).

Le *cultivar* di frumento oggi coltivate derivano infatti dalla domesticazione dei progenitori selvatici del frumento e appartengono essenzialmente a due specie biologiche: il grano tenero esaploide, *Triticum aestivum* ($2n = 6x = 42$ cromosomi), utilizzato principalmente per la produzione di farine destinate alla panificazione; e il grano duro tetraploide, *T. turgidum subsp. durum* ($2n = 4x = 28$ cromosomi), utilizzato principalmente per la produzione di pasta, semola e cous cous (Peng, Sun, e Nevo 2011b). Le altre specie appartenenti al genere *Triticum* hanno avuto un ruolo estremamente

importante nel passato, anche se oggi non sono più coltivate in quanto non risultano economicamente vantaggiose. Possono essere definite relitti, che spesso si ritrovano solo nella forma selvatica che cresce spontaneamente nei luoghi di origine (Zohary e Hopf 2000; Gill *et al.* 2006).

In termini più generali, al genere *Triticum spp.* appartengono diverse specie, selvatiche, coltivate, o in entrambe le forme. Queste specie vengono generalmente classificate secondo il loro livello di ploidia: esistono infatti frumenti diploidi, tetraploidi ed esaploidi (Feldmann *et al.* 1995). Secondo la classificazione odierna il genere *Triticum* è costituito da sei specie differenti: *Triticum monococcum* L.; *Triticum urartu* Tumanian ex Gandilyan; *Triticum turgidum* L.; *Triticum timopheevii* (Zhuk.); *Triticum aestivum* L.; e *Triticum zhukovskyi* Menabde & Ericz. Queste specie sono raggruppate in tre sezioni (Matsuoka 2011):

- 1) Monococcum, che comprende tutti i frumenti diploidi. A questa prima sezione appartengono due specie, la prima detta anche piccolo farro (genoma A^mA^m), che si divide ulteriormente in altre due sottospecie: *T. monococcum aegilopoides*, selvatico, e *T. monococcum monococcum*, coltivato. La seconda specie è *T. urartu* (genoma A^uA^u), la quale esiste solo nella forma selvatica. Nonostante i genomi di *T. monococcum* e *T. urartu* siano molto simili, questi due frumenti sono considerati due specie distinte poiché se interfecondate non danno progenie fertile (Zohary e Hopf 2000). A queste due specie se ne deve aggiungere una terza, portatrice del genoma BB. Si ritiene che il genoma B derivi da un antenato prossimo della specie *Aegilops speltoides*, portatore del genoma GG (Huang *et al.* 2002; Dvorak e Akhunov 2005). Nonostante questa specie non sia stata

identificata con assoluta certezza in tempi moderni, ha un'importanza fondamentale nella comparsa dei frumenti tetraploidi ed esaploidi;

- 2) Dicoccoidea, nella quale sono raccolte le specie tetraploidi. Anche in questo gruppo si distinguono due rami che portano a due specie diverse: il primo è quello di *T. turgidum* (ibridizzazione tra *T. uratu* e specie affine a *Ae. speltoides*, genoma A^uA^uBB), che include numerose sottospecie. Tra queste svariate sottospecie quella che per l'umanità è di maggiore interesse è *T. turgidum subsp. dicoccum*. Da questo cereale, chiamato comunemente farro, deriva il *T. turgidum subsp. durum* cioè il grano duro moderno. Il secondo ramo porta a *T. timopheevii* (genoma A^uA^uGG): sebbene questa specie risulti molto simile a *T. turgidum*, essa è il risultato di una più recente ibridizzazione tra *Aegilops speltoides* e il polline di *T. urartu*. A *T. timopheevii* appartengono due sottospecie, la prima è *T. timopheevii armeniacum*, la forma selvatica la cui domesticazione ha dato origine alla seconda sottospecie *T. timopheevii timopheevii*, coltivata in alcune regioni del Caucaso (Zohary e Hopf 2000);
- 3) Triticum, in cui troviamo le specie esaploidi *T. aestivum* (genoma A^uA^uBBDD), il frumento tenero moderno, e *T. zhukovskyi* (genoma A^mA^mA^uA^uGG), che deriva dall'ibridizzazione tra *T. timopheevii timopheevii* e il polline di *T. monococcum*. *Triticum zhukovskyi* si trova solo nella forma coltivata ed è presente in alcune regioni del Caucaso, dove ricopre un'importanza limitata (Zohary e Hopf 2000).

La classificazione dei frumenti precedentemente descritti è riportata in Tabella 1, in cui viene evidenziata la divisione delle varie specie e sottospecie che compongono il genere *Triticum*.

Tabella 1. Divisione delle specie e sottospecie appartenenti al genere *Triticum* (dopo Van Slageren 1994). Nella tabella è anche riportata la composizione del genoma e se conosciuto il nome comune del frumento.

Section	Species and subspecies	Genome constitution	Examples of common names ^b	
Monococcon	<i>Triticum monococcum</i> L.	AA		
	subsp. <i>aegilopoides</i> (Link) Thell.		Wild einkorn	
	subsp. <i>monococcum</i>		Cultivated einkorn	
Dicoccoidea	<i>Triticum urartu</i> Tumanian ex Gandilyan	AA		
	<i>Triticum turgidum</i> L.	AABB		
	subsp. <i>dicoccoides</i> (Körn. ex Asch. & Graebn.)		Wild emmer	
	subsp. <i>dicoccon</i> (Schrank) Thell.		Cultivated emmer	
	subsp. <i>durum</i> (Desf.) Husn.		Durum or macaroni wheat	
	subsp. <i>polonicum</i> (L.) Thell.		Polish wheat	
	subsp. <i>turanicum</i> (Jakubz.) Á.Löve & D.Löve		Khorassan wheat	
	subsp. <i>turgidum</i>		Rivet wheat	
	subsp. <i>carthlicum</i> (Nevski) Á.Löve & D.Löve		Persian wheat	
	subsp. <i>paleocalchicum</i> (Menabde) Á.Löve & D.Löve		Georgian wheat	
	<i>Triticum timopheevii</i> (Zhuk.) Zhuk.		AAGG	
	subsp. <i>armeniicum</i> (Jakubz.) van Slageren			Wild timopheevii
	subsp. <i>timopheevii</i>			Cultivated timopheevii
Triticum	<i>Triticum aestivum</i> L.	AABBDD		
	subsp. <i>aestivum</i>		Common wheat	
	subsp. <i>compactum</i> (Host) MacKey		Bread wheat	
	subsp. <i>sphaerococcum</i> (Percival) MacKey		Club wheat	
	subsp. <i>macha</i> (Dekapr. & Manabde) MacKey		Indian dwarf wheat	
	subsp. <i>spelta</i> (L.) Thell.		Spelt	
	<i>Triticum zhukovskyi</i> Menabde & Ericz.	AAAAGG		

Da un punto di vista genetico, il grano moderno è il risultato della fusione di tre genomi omeologhi, ovvero derivanti da specie evolutivamente correlate. A differenza di cromosomi omologhi, i cromosomi omeologhi sono solo parzialmente uguali, e questo pone delle limitazioni al loro accoppiamento e ricombinazione. Questa unione si è verificata in seguito all'ibridizzazione spontanea tra alcune specie di frumento, portatrici del genoma A, B e D (Figura 3).

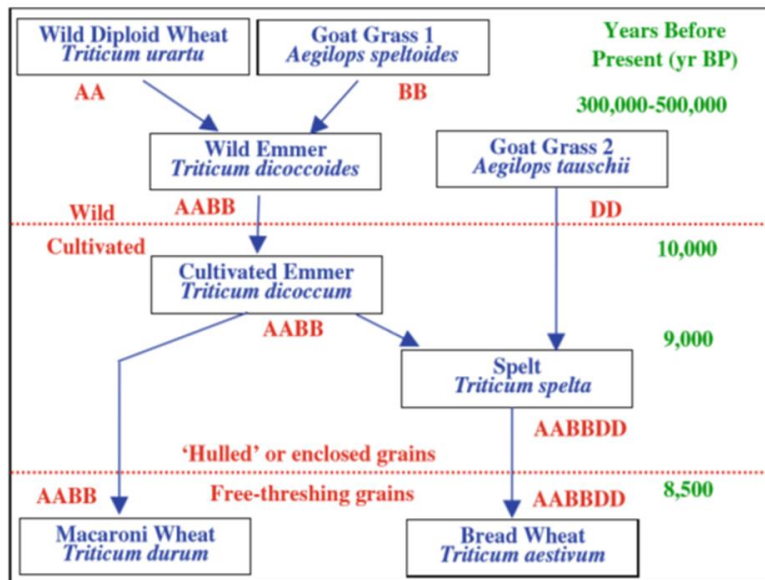


Figura 3. L'evoluzione del frumento dalle specie selvatiche dell'Età della Pietra fino ai frumenti moderni (<http://www.newhallmill.org.uk/wht-evol.htm>).

La storia evolutiva del frumento coltivato inizia con l'ibridizzazione tra due specie diploidi selvatiche, il donatore del genoma A^u *T. urartu* (Figura 4a) ed il donatore del genoma B, una specie appartenente alla linea evolutiva della odierna specie selvatica *Aegilops speltoides* ($2n = 2x = 14$, genoma SS, Figura 4b) (Kilian *et al.* 2007). L'unione tra queste due specie ha portato alla formazione del tetraploide selvatico *T. turgidum ssp. dicoccoides* ($2n = 4x = 28$, genoma A^uA^uBB, Figura 4d) (Huang *et al.* 2002; Dvorak e Akhunov 2005). Circa 10,000 anni fa, con l'avvento dell'agricoltura nella Mezzaluna Fertile, l'uomo iniziò a coltivare il farro selvatico (Salamini *et al.* 2002) esercitando inconsciamente una forte selezione direzionale per caratteri di interesse. Questa selezione portò in seguito alla formazione di *T. dicoccum* ($2n = 4x = 28$, genoma A^uA^uBB), la forma coltivata del farro. Quest'ultimo si ibridizzò spontaneamente con il diploide selvatico *Ae. tauschii* ($2n = 2x = 14$, genoma DD, Figura 4c) per produrre *T. spelta* ($2n = 6x = 42$, genoma A^uA^uBBDD) (Matsuoka e Nasuda 2004). Quest'incrocio ha avuto luogo nelle

regioni a sud del Mar Caspio, in seguito all'espansione di *T. dicoccum* dalla Mezzaluna Fertile all'habitat naturale di *Ae. Tauschii* (Salamini *et al.* 2002; Giles e Brown 2006).

Una mutazione naturale su di *T. dicoccum* e *T. spelta* trasformò poi le spighe di queste due specie da *hard threshing*, cioè con glume coriacee e resistenti e semi rivestiti, a spighe più semplici da raccogliere definite come *free-threshing*, che presentano cioè glume delicate e semi non rivestiti (Peng, Sun, e Nevo 2011b). La mutazione che ha portato al cambiamento nel fenotipo *free-threshing* delle spighe è stato un'ulteriore passo nella storia evolutiva del frumento verso *T. turgidum ssp. durum* (Figura 4e) e *T. aestivum* (Figura 4f), le due principale specie di frumento oggi coltivate.



Figura 4. Le spighe di grano appartenute alle forme selvatiche e domesticate dei frumenti. Progenitori diploidi selvatici dei frumenti tetraploidi e esaploidi: (a) *Triticum urartu* (A^uA^u), (b) *Aegilops speltoides* (SS) e (c) *Aegilops tauschii* (DD). Forme selvatiche e domesticate dei frumenti tetraploidi: (d) *Triticum turgidum ssp. dicoccoides* (A^uA^uBB) e (e) *Triticum turgidum ssp. durum* (A^uA^uBB). Il grano tenero coltivato *Triticum aestivum* (A^uA^uBBDD) (f), che si è originato in seguito all'ibridizzazione tra la forma domesticata di *T. turgidum ssp. dicoccum* e il diploide selvatico *Aegilops tauschii*, modificata da Feuillet *et al.*, 2008.

1.2 Sfide dell'agricoltura moderna

L'adattamento dell'agricoltura ai cambiamenti climatici globali è una delle sfide più urgenti e importanti dei nostri giorni. Le colture devono infatti continuare a fornire le

risorse necessarie a garantire la sopravvivenza dell'umanità resistendo a nuove pressioni evolutive introdotte dalle modificazioni del clima (CWR e Climate Change, 2013). Questa sfida è resa più ardua dal fatto che accoppiato al cambiamento climatico globale troviamo un'ulteriore problema, ovvero l'aumento della popolazione. Già nel 2001 la FAO indicava che la popolazione mondiale sarebbe arrivata a 8,3 miliardi nel 2030 e a 9,3 miliardi nel 2050; attualmente ammonta a oltre 6 miliardi (<http://www.fao.org>). La crescita demografica ha determinato un aumento nell'urbanizzazione, ed il concomitante miglioramento della qualità della vita ha portato i consumatori a preferire alimenti più ricchi ed elaborati come la pasta e la carne e ad allontanarsi invece da alimenti considerati poveri come il riso e le patate (Gustafson *et al.* 2009a). Questo cambiamento nelle abitudini alimentari è tanto più evidente nei paesi in via di sviluppo a rapida crescita economica. Stime effettuate a partire di questi dati prevedono un netto aumento della domanda mondiale di cibo, che richiederà un incremento di circa il 66% della produzione agricola entro il 2040 (Gustafson *et al.* 2009a). Allo stesso tempo, è pressoché impossibile destinare nuovi ettari di terreno alla coltivazione (Young 1999): l'agricoltura si trova quindi ad affrontare l'insieme di queste nuove sfide in un sistema chiuso.

La prerogativa all'adattamento delle risorse agronomiche a nuove condizioni ambientali è la loro plasticità e resilienza. La diversità genetica è un elemento fondamentale e necessario che fornisce a una pianta la capacità di resistere ai vari tipi di *stress*. Può essere considerata come il materiale grezzo per l'adattamento, in quanto la disponibilità di geni ed alleli favorevoli, e le loro combinazioni, determinano la possibilità di una popolazione vegetale a occupare una determinata nicchia ecologica. Una pianta sopravvivrà a determinate condizioni ambientali solo se è in possesso di combinazioni alleliche che le permettono di crescere in quella condizione. La variabilità allelica che

osserviamo sul campo, nelle popolazioni naturali, è dunque il frutto dell'evoluzione, con le sue numerose spinte tra cui la selezione naturale. Le linee coltivate, ottimizzate per la resa e l'uniformità, non sono plasmate dalla selezione naturale, risultando più vulnerabili agli *stress* biotici e abiotici. L'agricoltura moderna deve fare i conti con la drastica riduzione della diversità genetica nelle linee coltivate, contribuita dai programmi di miglioramento genetico effettuati negli anni (Haudry *et al.* 2007).

La domesticazione di una pianta è infatti una vera e propria selezione, operata dall'uomo, su determinati caratteri agronomici allo scopo di rendere le piante più semplici da coltivare, da raccogliere e soprattutto più produttive. Questo processo di selezione rende le piante dipendenti dall'uomo e culmina in ciò che è stata definita 'sindrome da domesticazione' (Hawkes 2013; Zohary e Hopf 2000; Fuller 2007; Harlan, Wet, e Price 1973). Per quanto riguarda i cereali, la sindrome da domesticazione si compone di sette caratteri principali (Brown *et al.* 2009; Fuller 2007):

1. disseminazione dei semi. Le specie domesticate presentano una mutazione che rende il rachide più tenace, permettendo in questo modo al chicco di grano di rimanere attaccato alla spiga matura. Questa caratteristica è molto importante perché rende la propagazione della pianta dipendente dall'uomo, fornendo all'agricoltore rendimenti più elevati in quanto la raccolta può essere ritardata fino a completa maturità dei chicchi (Hillman e Davies 1990; Bar-Yosef 1998);
2. organi di dispersione. Perdita di organi come tricomi, ganci e reste, che facilitano la dispersione anemocora e zoocora dei semi. L'uomo ha adoperato questa selezione per favorire i processi di pulizia del chicco in post-raccolta;

3. aumento della taglia dei semi. Questo carattere è spesso identificato come segno identificativo dell'intervento dell'uomo, in quanto l'agricoltore opera questa selezione per incrementare le rese produttive (Harlan, Wet, e Price 1973);
4. ridotta sensibilità verso l'ambiente. Il frumento, così come molte altre piante, risponde agli stimoli ambientali; quindi come accade normalmente in natura, un seme germinerà solo in condizioni favorevoli. Le specie coltivate hanno perso questa caratteristica, mostrando invece una rapida germinazione dopo la semina;
5. sincronizzazione nel periodo di fioritura. Questo carattere consente di evitare la raccolta scalare dovuta alle diverse epoche di fioritura e quindi ridurre tempi e costi;
6. portamento compatto. Al fine di rendere la raccolta più semplice;
7. alterazione dei valori nutrizionali. Cercare di migliorare le caratteristiche organolettiche della farina, come per esempio aumentare il contenuto di proteine per aumentare la forza di panificazione e quindi migliorare la qualità del pane (Brown 1999).

Durante questo processo di domesticazione e selezione, le linee coltivate hanno subito diversi colli di bottiglia causati dalla selezione intergenerazionale delle linee più favorevoli operata dai coltivatori. Questo ha portato a una drastica riduzione della diversità genetica che sottende alla variabilità fenotipica su cui la selezione è stata operata.

Con questi presupposti, la situazione futura delle risorse alimentari sembra essere abbastanza critica; se la domanda mondiale è in continuo aumento, le nostre scorte di frumento sono in drastica diminuzione (Gustafson *et al.* 2009) ed esposte come mai a

rischi biotici ed abiotici. L'ampliamento delle basi genetiche dell'adattamento delle specie coltivate è una delle possibili soluzioni alla limitatezza di risorse che si prevede negli anni a venire. I progenitori selvatici, che possiedono un'elevata diversità genetica non intaccata dalla selezione umana, possono essere utilizzati per obliterare il collo di bottiglia sofferto dalle principali *cultivar* durante domesticazione e miglioramento. Con le tecnologie moderne è possibile riavvolgere il processo di domesticazione e ripensare drasticamente alla variabilità adattativa necessaria nelle colture odierne. Il possibile impiego e l'utilità di questi progenitori delle specie coltivate nei programmi di miglioramento genetico fu riconosciuta già negli anni 1940 e 1950 (Plucknett *et al.* 1987). In seguito, il loro utilizzo prese sempre più campo, portando a considerare un numero sempre più ampio di specie di piante (Hoyt 1988). Già Tanksley e McCouch (1997) sottolinearono l'importanza del mappaggio e lo sfruttamento genetico dei genomi selvatici, suggerendone un continuo campionamento, al fine di incrementare le possibilità di scoperta di nuovi geni utilizzabili nel miglioramento genetico.

Le specie selvatiche hanno come vantaggio di non essere andate incontro al *bottleneck* genetico generato dalla domesticazione, a cui invece sono state soggette le linee coltivate. I progenitori selvatici sono il frutto della selezione naturale, possiedono cioè quei caratteri di rusticità che li rendono resistenti e resilienti a vari tipi di *stress* (siccità, alte temperature, malattie, ecc.), e che rendono possibile l'adattamento di queste piante alle differenti condizioni ambientali che incontrano nel loro areale di distribuzione (Jackson, Ford-Lloyd, e Parry 2013a). Metodi molecolari sviluppati negli ultimi decenni permettono di avere una caratterizzazione fine della diversità molecolare dei progenitori selvatici. I caratteri di interesse agronomico sono spesso il risultato dell'interazione di numerosi loci genici, che di concerto determinano i fenotipi complessi: questi sono detti

quantitative trait loci (QTL), poiché si manifestano in maniere quantitativa e non qualitativa come i più classici caratteri Mendeliani. La capacità di mappare e caratterizzare tali loci genomici nei progenitori selvatici permette in ultima istanza di identificare i geni che determinano il fenotipo di interesse, e da qui di sfruttare la presenza di alleli “selvatici” favorevoli che sono stati da tempo perduti nel processo di miglioramento del frumento di *elite*.

Un metodo diretto per sfruttare al meglio questi nuovi caratteri di rusticità è quello di incrociare le linee *elite* coltivate con le specie selvatiche. Il punto di partenza è dato dalla disponibilità di nuova variabilità allelica nei progenitori selvatici che crescono attualmente in differenti condizioni ambientali. Nel frumento è possibile sfruttare la compatibilità genetica dei progenitori con la loro controparte coltivata in incroci che mimano il processo evolutivo del frumento moderno. Le diverse sottospecie di frumento presentano infatti genomi omeologhi che permettono una più semplice identificazione ed un eventuale trasferimento della omeologa variazione allelica in genomi di interesse, quali quelli del frumento duro e tenero. La diversità genetica dei progenitori potrebbe quindi integrarsi e completare i meccanismi pre-esistenti di resistenza allo *stress* nel frumento coltivato (Reynolds, Dreccer, e Trethowan 2007). Questa strategia è stata applicata numerose volte nell’ambito dei frumenti, portando a dei risultati di grande rilevanza agronomica. Uauy *et al.* (2005) ha identificato nel genoma di *T. dicoccoides* *Yr36*, un gene coinvolto nella resistenza alla ruggine, questo gene è stato successivamente trasferito e clonato nel grano duro (Fu *et al.* 2009). Qiu *et al.* (2005) è riuscito a mappare, sul genoma di *T. urartu* utilizzando dei microsatelliti, il gene di resistenza all’oidio e poi trasferire quest’ultimo su grano tenero. Ulteriori esempi sono forniti dall’identificazione del gene di resistenza (*Hdic*) alla mosca del grano (*Mayetiola destructor*. Say) scoperto

in *T. dicoccum* (Liu *et al.* 2005), e del gene di resistenza alla ruggine gialla (*Yr26*) scoperto in *T. turgidum* (Ma *et al.* 2001); entrambi sono stati poi trasferiti in *T. aestivum*.

Esistono tre metodi di trasferimento dell'informazione proveniente dai progenitori selvatici al frumento coltivato:

1. introgressione diretta, cioè il movimento di geni da una specie ad un'altra attraverso ripetuti *backcrossing* tra l'ibrido interspecifico e il suo parentale, frumento in questo caso. In questa metodica si selezionano tipicamente uno o più genotipi selvatici di interesse (ad esempio, per resistenza ad un determinato patogeno) e si procede a trasportare le regioni genomiche responsabili delle caratteristiche desiderabili nel *background* di un genoma adattato all'agricoltura tramite incrocio (Figura 5). L'uso di marcatori molecolari permette di tracciare l'introgressione del locus di interesse e di monitorarne la presenza nel genoma *target*. Ripetuti *backcrossing* con il parentale coltivato sono necessari per "ripulire" il genoma. Spesso, infatti, le specie selvatiche possiedono oltre che al carattere di rusticità di interesse, numerosi altri caratteri che non sono adatti all'agricoltura (ad esempio la disarticolazione delle spighe dal rachide e successiva disseminazione dei semi, bassa produttività, ecc.). I marcatori molecolari permettono di seguire e tenere traccia del gene di interesse durante i ripetuti *backcrossing* e selezionare quindi solo quelle linee in cui esso è contenuto, nel contempo rimuovendo tutte le regioni indesiderate. Al termine del processo, si ottengono linee *inbreed* aventi il gene di resistenza selvatico in un *background* genomico domestico. Questa tecnica è economica, non necessita di particolari

tecnologie, e quindi può essere effettuata in molti laboratori di ricerca. Inoltre, è una metodologia largamente accettata dalla popolazione e questo rende il suo utilizzo immediato. Il principale svantaggio invece è che risulta essere poco efficiente, questo perché queste introgressioni interspecifiche (ovvero tra specie diverse) presentano una bassa fertilità e scarse ricombinazioni (Fricano *et al.* 2014), rendendo difficile dapprima l'introggressione del carattere di interesse, e in seguito la pulizia del genoma tramite i ripetuti *backcross*.

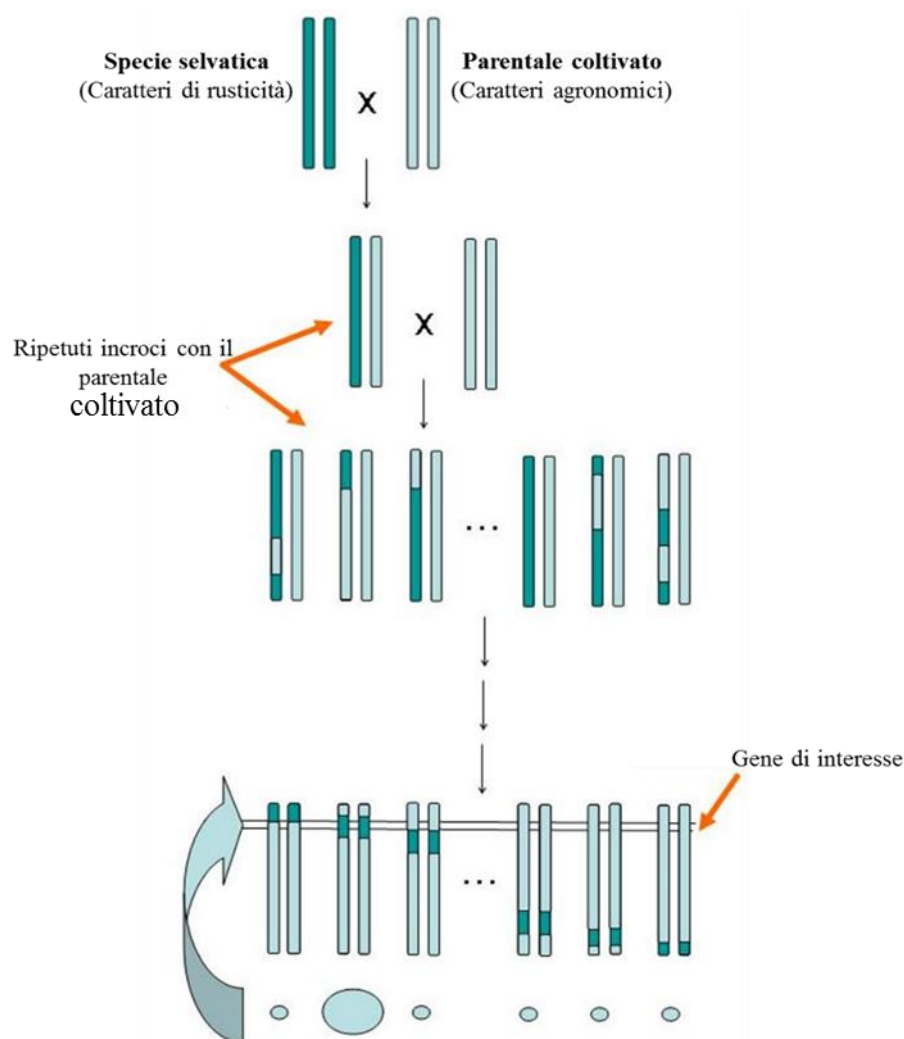


Figura 5. Immagine che illustra la tecnica di introgressione interspecifica. La specie selvatica porta i caratteri di rusticità mentre il parentale coltivato i caratteri agronomici favorevoli. I marcatori molecolari vengono utilizzati per monitorare la presenza del locus di interesse (la cui posizione è nota) attraverso le successive fasi di *backcross* (Vadez *et al.* 2012).

2. *targeting* delle sequenze omologhe, che si basa sull'utilizzo delle specie selvatiche per identificare il gene di interesse. Siccome tali specie presentano la variabilità necessaria a condurre mappaggio di *QTL*, è possibile utilizzarle per mappare un determinato locus genico. Una volta mappato, la sequenza del gene viene poi ricercata nel frumento per omologia con quella del progenitore

selvatico. A questo punto è possibile effettuare uno screening per la presenza dell'allele favorevole nelle collezioni di frumenti coltivati (che contano migliaia di varietà dalle così dette *landraces*, o materiale primitivo, alle moderne *cultivar*), osservando il solo locus di interesse. L'identificazione dell'allele favorevole in una determinata varietà ne permette il trasferimento con alta efficienza in linee *elite*, utilizzando incroci intra-specifici. Il trasferimento del carattere risulta essere in questo caso molto più semplice rispetto a quello con un'altra specie, poiché non sono presenti dei problemi di incompatibilità e bassa fertilità. A limitare questo approccio c'è il possibile cambio di ruolo del locus selvatico nella fisiologia della pianta coltivata: questo può far sì che il gene risulti il medesimo per omologia di sequenza, ma che la funzione nel frumento coltivato sia in qualche misura diversa da quella coperta nei progenitori selvatici. Inoltre, è possibile che la variante allelica di interesse nelle specie selvatiche non sia più presente nelle varietà di frumento coltivato, rendendo necessario ricorrere al caso 1 o al caso 3.

3. trasformazione, che consente di prelevare il gene di interesse dalla specie selvatica e poi inserire direttamente il DNA esogeno nel genoma del frumento coltivato. Dato che viene selezionato un determinato gene, il trasferimento risulta "pulito" cioè non saranno presenti eventuali caratteri non desiderati, come avviene invece nell'introggressione diretta. Il problema principale di questo approccio è che richiede il clonaggio del gene e quindi studi più lunghi e dispendiosi. Un'ulteriore problema sorge dal fatto che gli organismi geneticamente modificati (OGM) non sono sempre accettati dalla popolazione e per questo motivo di difficile utilizzo. A livello europeo questo tema ha

portato a numerose controversie. Per diversi anni infatti le normative che disciplinano gli OGM, una direttiva del 2001 (Direttiva 2001/18/CE) e due Regolamenti del 2003 (1829 e 1830/2003/CE), sono state oggetto di discussioni, lasciando un vuoto legislativo che non ha consentito la regolare coltivazione delle piante geneticamente modificate (GM), nonostante l'autorizzazione della comunità Europea, in alcuni Stati membri come l'Italia. La direttiva (UE) 2015/412 approvata il 13 gennaio, mira a modificare e compensare le altre due normative precedenti in modo da favorire la coltivazione di piante GM sul suolo Europeo (www.europarl.europa.eu).

1.3 Il progenitore selvatico *Triticum urartu*

Uno dei più importanti progenitori selvatici del frumento è *T. urartu*, il donatore del genoma A^u di *T. turgidum subsp durum* e *T. aestivum*. *Triticum urartu* Thum. ex Ghandil (Figura 6) appartenente alla famiglia delle Poaceae genere *Triticum*, è una pianta erbacea annuale con radice fascicolata. Il fusto è un culmo costituito da nodi e internodi lungo circa 60-90 cm, generalmente in corrispondenza degli internodi risulta essere cavo. Le foglie, che si inseriscono a livello dei nodi pubescenti (caratteristica delle specie selvatiche), sono costituite da una guaina avvolgente il culmo stesso. Le foglie presentano una forma lanceolata e nervature parallele. Ogni culmo porta all'apice l'infiorescenza, composta da racemi, chiamata spiga. Ogni spiga è formata da un rachide fragile, costituito da internodi oblungi ognuno portante una spighetta; a maturità queste si disarticolano dal rachide e cadono sul terreno. Le spighette sono oblunghe e compresse lateralmente, ognuna è formata da due glume, di cui l'inferiore prende il nome di lemma che si presenta coriacea, dentata all'apice e con arista (o resta), e la superiore, più piccola e generalmente

racchiusa dalla precedente, viene detta palea. I fiori che si trovano all'interno delle glume non sono tutti fertili, solitamente questi si trovano soprattutto nella zona centrale della spiga mentre quelli basali e apicali risultano essere sterili o comunque poco sviluppati. Il frutto detto comunemente granella o chicco è una cariosside, un frutto secco indeiscente dalla forma ovale (<http://www.kew.org/data/grasses-db.html>).



Figura 6. Immagine che rappresenta *Triticum urartu* Thum. ex Ghandil. (www.agbina.com/site.xp/053051056124049054053050.html).

Triticum urartu cresce nella Mezzaluna Fertile in diverse condizioni pedoclimatiche, in *range* differenti di altitudine, temperatura e piovosità. Popolazioni

naturali di *T. urartu* non hanno subito processi di domesticazione (Feldman e Levy 2005) e pertanto non sono andate incontro ai *bottleneck* genetici delle specie coltivate. Le popolazioni naturali di *T. urartu* quindi presentano variabilità allelica unicamente contribuita dai processi evolutivi, tra cui la selezione naturale. Sebbene da un punto di vista agronomico *T. urartu* non risulti particolarmente interessante (bassa produttività, disseminazione dei semi, ed altri caratteri primitivi), il suo genoma selvatico presenta proprio caratteri di rusticità apportati dalla selezione naturale (Rouse e Jin 2011), che potrebbero essere utilizzati nei programmi di miglioramento genetico del frumento. La presenza di geni coinvolti nella tolleranza a stress biotici in *T. urartu* è stata già dimostrata negli studi di Rouse e Jin (2011b) e di Qiu *et al.* (2005) effettuati per la resistenza a due malattie fungine: *Puccinia graminis f. sp. tritici* e *Blumeria graminis f. sp. tritici*, rispettivamente. *T. urartu* ha dimostrato una buona tolleranza a *stress* di natura abiotica, come la siccità, la presenza eccessiva di sale e le alte temperature (Valkoun 2001; Kara *et al.* 2000; Gorham *et al.* 1991).

La ricerca di questi caratteri in *T. urartu* è oggi resa più efficiente dal recente sequenziamento del suo genoma (Ling *et al.* 2013), che con 4,94 miliardi di basi rappresenta la metà del genoma di frumento duro ed un terzo del genoma di frumento tenero. I cromosomi di *T. urartu* sono omeologhi ai cromosomi del genoma di grano duro e tenero, e questo consente un trasferimento efficiente di alleli di interesse da questi "cromosomi selvatici" a quelli coltivati (Qiu *et al.* 2005). Per queste ragioni *T. urartu* è considerato una risorsa fondamentale per il miglioramento genetico del frumento (Henry e Nevo 2004).

1.4 Strategie di indagine della diversità di *Triticum urartu*

Le vaste collezioni di *T. urartu* raccolte attraverso la Mezzaluna Fertile rappresentano quindi un'importante risorsa per il miglioramento di frumento, sia per la similarità di *T. urartu* con il frumento coltivato, sia per la disponibilità di strumenti molecolari atti alla caratterizzazione della sua diversità. In questa cornice, la *landscape genomics* è una pratica che può essere utilizzata per usufruire al meglio delle popolazioni naturali di *T. urartu* per identificare loci genici deputati all'adattamento. La *landscape genomics* analizza la distribuzione spaziale degli individui di popolazioni naturali, ed unendo tecniche geografiche, di genetica di popolazione e tecniche statistiche derivate dal mappaggio di *QTL*, si propone di scoprire loci genici coinvolti nell'adattamento (Luikart *et al.* 2003). Questa pratica nasce da una considerazione solo in apparenza banale: se una popolazione di piante cresce sottoposta a determinate condizioni ambientali, essa possiederà la variabilità allelica necessaria a sopravvivere in quell'ambiente. Per esempio, se una pianta cresce in un'area a bassa piovosità si può supporre possieda i meccanismi molecolari che gli consentono di sopravvivere a *stress* idrici. Ugualmente, popolazioni che prosperano in presenza di patogeni possiederanno meccanismi di resistenza come risultato della selezione naturale. La *landscape genomics* è quindi il risultato della fusione tra due tipi di informazione, i primi ottenuti tramite analisi molecolari, mentre i secondi da informazioni geografiche (Turner, Gardner, e O'Neill 2001) (Figura 7).

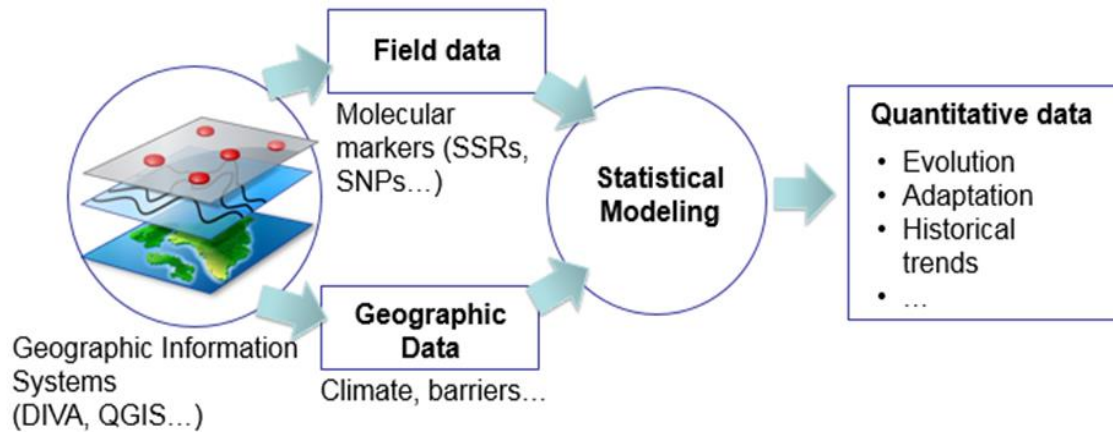


Figura 7. *Landscape genomics*, una disciplina che unisce i dati molecolari a quelli geografici. Un sistema geografico informatico è utilizzato per rappresentare l'area geografica di interesse, e studiare il rapporto spaziale tra dati di campo (accessioni campionate e loro diversità genetica) e dati geografici (clima, suolo). La coincidenza spaziale tra queste quantità viene analizzata con indagini statistiche capaci di metterle in relazione, e rispondere in maniere quantitativa a domande sull'evoluzione, l'adattamento, ed altro.

Tramite i sistemi di informazioni geografica (GIS, da *geographic information systems*) è possibile ottenere i dati geografici necessari alla componente paesaggistica (il *landscape*) della *landscape genomics*. Il GIS è un sistema che permette la visualizzazione e l'analisi di oggetti geografici, ovvero elementi relativi a determinate posizioni sulla superficie terrestre – una sorta di cartografia informatica. In un sistema GIS è possibile mettere in relazione più dati associati a una coordinata geografica di modo da comprendere le relazioni che intercorrono tra i diversi elementi. È possibile per esempio caricare sul programma la mappa geografica di una regione e, a questa, sovrapporre altre informazioni come i dati ambientali (piovosità, temperatura, altitudine, ecc.) relativi a quella zona. Se a questa rappresentazione multi-strato aggiungiamo i siti di campionamento di popolazioni di piante di interesse, ottenuti tramite comuni ricevitori *Global Positioning System* (GPS), è possibile trasferire le informazioni geografiche su

questi punti, e quindi associare a ciascuna popolazione campionata misure quantitative riferite al clima, al suolo, all'antropizzazione e via dicendo.

La componente genomica della *landscape genomics* si riferisce invece all'analisi della variabilità genetica contenuta negli individui campionati. Allo scopo di identificare loci genici deputati all'adattamento, la *landscape genomics* necessita di una grande densità di marcatori molecolari che siano in grado di fornire una descrizione *genome-wide* della specie di interesse. Negli ultimi anni si è assistito a un notevole sviluppo dei marcatori molecolari, dapprima passati da metodi di *genotyping* a bassa densità ed alta informatività, quali i microsatelliti (SSR, *Single Sequence Repeat*), a metodi ad alta densità ma bassa informatività quali gli AFLP (*Amplified Fragment Length Polymorphism*). Gli SSR sono dei marcatori codominanti, permettono infatti di discriminare la variabilità presente all'interno del microsatellite e quindi le varianti alleliche di un determinato locus genico. Forniscono dunque un alto livello di informazione, ma solo per un determinato locus genico, e solo limitati nel numero dalla laboriosità della genotipizzazione (Gupta *et al.* 1996; Schlötterer e Pemberton 1998). Gli AFLP sono invece dominati e multilocici, in quanto il metodo AFLP si basa sull'utilizzo di enzimi di restrizione e l'osservazione dei frammenti ottenuti al fine di ottenere un numero elevato di marcatori molecolari sparsi per il genoma. In questo modo è possibile avere un'informazione generale delle variabilità presente tra differenti individui, ma non è possibile discriminare la variabilità allelica di un determinato locus genico, poiché l'osservazione di frammenti di restrizione permette di determinare la sola presenza o assenza di una sequenza di taglio (Vos *et al.* 1995). Entrambi questi metodi sono stati usati nella *landscape genetics* (Storfer *et al.* 2010; Hammami *et al.* 2014). Recentemente, i nuovi metodi di sequenziamento basati sull'analisi di sequenza di frammenti di

restrizione analoghi a quelli prodotti tramite tecnica AFLP hanno permesso di passare dalla *landscape genetics* alla *landscape genomics*, ottenendo un'alta informatività associata ad una rappresentazione diffusa del genoma. L'applicazione di tecniche come il *restriction-site associated DNA sequencing* (RADSeq), che producono un elevato numero di marcatori molecolari ad alta informatività (Davey e Blaxter 2010), e la sempre maggiore disponibilità di sequenze genomiche tra cui quella di *T. urartu* (Ling *et al.* 2013), rendono più semplice lo studio della variabilità allelica e quindi l'identificazione di loci di interesse agronomico in specie selvatiche. Il RADSeq si basa sull'utilizzo di enzimi di restrizione per tagliare il DNA in frammenti come nella tecnica AFLP, ma culmina nel sequenziamento dei frammenti prodotti. Per fare questo, è necessario associare a ciascun frammento prodotto sequenze identificative (*barcode*) che permettono di associare in maniera univoca una sequenza a un determinato individuo. I frammenti sono poi muniti di adattatori complementari alle sequenze che si trovano sulla *flowcell* della piattaforma Illumina, e viene così condotto il sequenziamento. Con questa tecnologia è possibile produrre decine di migliaia di SNP diffusi su tutto il genoma, e studiare l'adattamento genomico con una definizione ineguagliata. I dati molecolari ottenuti tramite RADseq sono fondamentali, perché sono in grado di fornire la più ampia rappresentazione della variabilità genetica all'interno della popolazione (Dell'Acqua *et al.* 2014).

La fusione dei dati geografici e quelli molecolari tramite modelli statistici, consente infine di studiare la distribuzione spaziale dei loci ecologicamente significativi e quindi facilitare la gestione delle risorse genetiche provenienti dall'adattamento delle piante all'ambiente (Manel *et al.* 2010). Collegando le informazioni geografiche con i dati ambientali è inoltre possibile mettere in evidenza la variabilità ambientale presente

in una popolazione e selezionare a priori individui che presentano determinate caratteristiche, per esempio piante che crescono in regimi a piovosità contrastante, e grazie a questo osservare solo i fenotipi delle piante campionate che presenteranno la variabilità allelica coinvolta in quello specifico fenotipo. Con questo tipo di unione è possibile massimizzare il risultato del *QTL mapping*, cioè l'utilizzo di regressioni lineari che mettono in relazione da un lato i genotipi con l'ambiente e dall'altro i genotipi con la diversità fenotipica.

2 Scopo della tesi

L'obiettivo della tesi è quello dell'iniziale caratterizzazione di una collezione di popolazioni naturali di *T. urartu*, donatore del genoma A del grano moderno, allo scopo di condurre uno studio di *landscape genomics* mirato alla scoperta di loci allelici rilevanti al miglioramento di frumento. Con lo scopo di osservare la più elevata variabilità genetica possibile, sono state aggregate 428 accessioni che rappresentano la reale distribuzione di *T. urartu* nella Mezzaluna Fertile. Le accessioni saranno caratterizzate ecologicamente, molecolarmente e fenotipicamente, allo scopo di fare convergere queste informazioni nell'identificazione di loci di interesse per il *breeding* di frumento. Nel corso di questa tesi, l'intera collezione è stata seminata in campo aperto allo scopo di amplificare il seme disponibile. Durante l'esperimento di campo, sono stati raccolti alcuni caratteri fenotipici in modo da fornire una prima caratterizzazione fenotipica della collezione. Allo stesso tempo, le informazioni geografiche sui siti di campionamento sono state proiettate in un sistema informativo geografico allo scopo di estrarre ed analizzare le caratteristiche ecologiche relative a ciascuna accessione. Il DNA genomico è stato estratto da tutte le accessioni e processato allo scopo di produrre dati di variazione molecolare. Il primo set di 24 campioni genotipizzati è stato analizzato per estrarre informazioni circa la diversità genetica inclusa nella collezione.

Il lavoro di tesi si è quindi articolato su tre livelli di caratterizzazione:

- geografica, studiando le variabili ambientali relative ai siti di campionamento dell'intera collezione, e quindi valutando le condizioni climatiche a cui ogni accessione è sottoposta;

- molecolare, utilizzando un protocollo di RADSeq in prima istanza su un *subset* di 24 individui. L'utilizzo di questa tecnica di *next-generation sequencing* ha come fine ultimo quello di scoprire la più elevata densità di polimorfismi a singolo nucleotide (SNP);
- fenotipica, valutando la distribuzione di diversi caratteri nella collezione di *T. urartu* cresciuta in campo.

3 Materiale e metodi

3.1 La collezione

Il materiale vegetale utilizzato in questa tesi è composto da diverse accessioni di *Triticum urartu* che rappresentano la più vasta collezione mondiale disponibile di questa specie. Le accessioni sono state ricercate nelle principali banche del germaplasma mondiali tramite il progetto Genesys (www.genesys-pgr.org), un portale che aggrega le collezioni di diverse istituzioni di conservazione e distribuzione. Una volta stilata una lista della accessioni univoche di *T. urartu*, queste sono state richieste sulla base della disponibilità, in particolare:

- i) 251 accessioni sono state ottenute dal *Germplasm Resources Information Network* in USA (USDA);
- ii) 68 accessioni sono state ottenute dal *National Institute of Agrobiological Sciences Genebank* (NIAS) in Giappone.

Le accessioni sono state fornite in sacchetti contenenti 5g di semi ciascuno. Per ciascuna accessione sono noti l'anno di campionamento, la località e, nella maggior parte dei casi, le coordinate geografiche esatte del sito di campionamento, in formato WGS84. Qualora le coordinate WGS84 non fossero presenti, era disponibile la località di campionamento sotto forma di *gazetteer*. Le coordinate geografiche mancati sono quindi state ottenute tramite *Google earth* (www.google.com/earth/), derivando le coordinate WGS84 a partire dalle località in cui le piante erano state campionate.

La collezione è stata integrata con 109 accessioni provenienti dal CRA - Istituto Sperimentale per la Cerealicoltura in Italia di Sant'Angelo Lodigiano (CRA-SAL),

fornite dal Dr. Andrea Brandolini. Queste accessioni sono in parte uniche ed in parte ridondanti con la collezione USDA. Le accessioni replicate, in quanto provenienti da diverse fonti, in questa fase sono state considerate distinte.

La collezione di 428 accessioni così ottenuta rappresenta la distribuzione di *T. urartu* attraverso tutta la Mezzaluna Fertile (Armenia, Iran, Iraq, Giordania, Libano, Siria e Turchia). Il numero delle accessioni campionate nei differenti paesi, e usate in questo studio, è riportato in *Tabella 2*. La collezione è stata mantenuta a 4°C fino al momento dell'utilizzo.

Tabella 2. Numero delle accessioni per paese di origine.

Paese	Numero delle accessioni
Armenia	20
Iran	18
Iraq	73
Giordania	6
Libano	110
Siria	45
Turchia	136
Sconosciuti	20

3.2 Analisi GIS

Le analisi GIS sono state condotte con due programmi GIS *open-source*: DIVA-GIS 7.5 (<http://www.diva-gis.org/>) e QGIS 2.10.1 (<http://www.qgis.org/>). Il primo è un sistema GIS appositamente sviluppato per la modellizzazione di nicchie ecologiche. Pur essendo meno efficiente e moderno di QGIS, per questa sua caratteristica è stata utilizzata per ottenere i dati ambientali dei siti di campionamento (anni~1950-2000, con precisione di ~5 Km) tramite *Worldclim* (Hijmans *et al.* 2005) sottoforma di variabili bioclimatiche (*Bioclim*). *Bioclim* è un *dataset* composto da un insieme di 19 variabili bioclimatiche ottenute da valori minimi e massimi di temperatura e precipitazioni mensili che

permettono di avere indicazioni sul clima di qualsivoglia posizione sulla superficie terrestre (Kriticos *et al.* 2012). Non essendo possibile registrare il clima in ogni punto della superficie terrestre, *Bioclim* è una superficie interpolata a partire dai dati misurati da una rete di stazioni climatiche dislocate in tutto il mondo. Le variabili bioclimatiche rappresentano diversi aspetti del clima che più influenzano la distribuzione delle specie vegetali, tra cui le tendenze annuali (temperatura media annua e precipitazione annuale), la stagionalità (range annuale di temperatura e precipitazioni), e i fattori limitanti alla crescita (come la temperatura del mese più freddo o più caldo) (Tabella 3).

I punti di campionamento sono stati importati in DIVA-GIS come *shapefile* vettoriale e sono stati utilizzati per estrarre l'altitudine e le variabili bioclimatiche in ciascun punto tramite la funzione "*extract climate by points*", ottenendo una matrice numerica di attributi bioclimatici da associare a ciascun campione. Tramite QGIS, le coordinate geografiche delle accessioni sono state importate come file vettoriale e ad esse sono state associate come attributi le variabili bioclimatiche ottenute da DIVA-GIS. QGIS è stato utilizzato per costruire mappe geografiche di altitudine, piovosità e temperatura a partire dai valori di *Bioclim* per queste tre variabili in formato *raster* per l'area di campionamento.

Tabella 3. Tabella in cui sono riportate i nome delle variabili bioclimatiche e il loro significato. In rosa, sono evidenziate le variabili bioclimatiche legate alla temperatura. In blu, le variabili bioclimatiche ascrivibili alla piovosità.

BIO₁	Temperatura media annuale
BIO₂	Media della variazione giornaliera (media del mese (T°max -T°min))
BIO₃	Isotermalità (BIO ₂ /BIO ₇) (*100)
BIO₄	Temperatura stagionale (deviazione standard * 100)
BIO₅	Temperatura max del mese più caldo
BIO₆	Temperatura min del mese più freddo
BIO₇	Variazione delle Temperatura annuale (BIO ₅ - BIO ₆)
BIO₈	Media della Temperatura del quartile più umido
BIO₉	Media della Temperatura del quartile più secco
BIO₁₀	Media della Temperatura del quartile più caldo
BIO₁₁	Media della Temperatura del quartile più freddo
BIO₁₂	Precipitazione annuale
BIO₁₃	Precipitazione del mese più umido
BIO₁₄	Precipitazione del mese più secco
BIO₁₅	Precipitazione stagionale (Coefficiente di variazione)
BIO₁₆	Precipitazione del quartile più umido
BIO₁₇	Precipitazione del quartile più secco
BIO₁₈	Precipitazione del quartile più caldo
BIO₁₉	Precipitazione del quartile più freddo

3.2.1 Gestione dati GIS

I dati relativi alla variabilità ecologica delle accessioni assemblati in QGIS sono stati esportati come file testuali ed importati in R (R Development Core Team 2015). R è un linguaggio di programmazione ad oggetti che consente di manipolare dati, effettuare analisi statistiche e rappresentare graficamente i risultati. R è un linguaggio ad alto livello basato su oggetti, ovvero utilizza un interprete che esegue passo dopo passo le istruzioni che gli sono state fornite. R è stato utilizzato per osservare la distribuzione delle accessioni in relazione alle variabili bioclimatiche relative ai siti di campionamento.

La rappresentazione grafica della distribuzione delle accessioni secondo le variabili bioclimatiche è stata ottenuta tramite la funzione *boxplot()* del pacchetto base R/*graphics*,

che permette una visualizzazione sintetica della distribuzione dei valori nei quattro quartili che la compongono. L'altitudine, pur non essendo inclusa nelle variabili bioclimatiche, è ragionevolmente alla base della variazione di molte di esse. La sua distribuzione nei campioni raccolti è stata quindi separatamente rappresentata tramite un istogramma con la funzione *hist()*, per mostrare in maggiore dettaglio la variazione altitudinale della collezione. Una volta osservata la distribuzione di tutte le variabili *Bioclim*, la funzione *cor()* ed il pacchetto R/corrplot (Wei 2013) sono stati usati per osservare il grado di correlazione tra i dati bioclimatici all'interno della collezione. In questo modo è stato possibile valutare quali variabili fossero correlate e potenzialmente ridondanti, e in quale misura.

La funzione *princomp()* di R è stata utilizzata per estrarre le componenti principali (PC) di una *principal component analysis* (PCA) (Mardia, Kent, e Bibby 1979; Venables e Ripley 2002) sui dati bioclimatici e sull'altitudine. La PCA è una tecnica matematica non parametrica che permette di ridurre la dimensionalità del dato, estraendo un numero limitato di PC ortogonali tra loro, ciascuna delle quali spieghi la massima parte della varianza del *dataset* originale. Tipicamente, le PC ottenute vengono ordinate per porzione di varianza spiegata, cosicché la prima PC è più significativa della seconda, la seconda PC lo è della terza, e così via. Le PC ottenute sono state utilizzate per visualizzare il dato ambientale in funzione della provenienza geografica dei campioni, e per valutare la presenza di *pattern* di variabilità a scala geografica ampia mantenendo una visione sintetica di insieme. Le PC sono state poi correlate con le variabili originali con il pacchetto R/corrplot (Wei 2013) per ricostruire il peso di ciascuna variabile originale su di esse: in questo modo, è possibile leggere il grafico della PCA alla luce del dato ecologico di partenza.

3.3 Germinazione del materiale ed estrazione del DNA

3.3.1 *Camera di crescita*

Due spighette sono state seminate per ciascuna accessione, in gruppi di 50 per volta, con lo scopo di ottenere il materiale vegetale verde necessario a condurre l'estrazione del DNA. I semi sono stati fatti germinare in vasi separati (\varnothing 4 cm) in una camera di crescita ad ambiente controllato (12h 24°C luce / 12h 22°C buio). L'irrigazione è stata effettuata secondo necessità, in modo da mantenere il terreno umido ma non troppo bagnato. Una volta iniziato l'accostimento, i tessuti verdi sono stati collezionati in egual misura da tutte le piantine germinate per ciascuna accessione separatamente, in misura di circa 150-200 mg di tessuto fresco ciascuna. Le piante sono state lasciate ricrescere ed è stata effettuata una seconda raccolta, allo stesso stadio fenologico, in modo da avere un *backup* per ogni accessione. Tutti i tessuti verdi sono stati inseriti in provette individuali da 1,5 ml ed etichettati. I campioni così raccolti sono stati conservati a -80°C fino al momento dell'estrazione del DNA.

3.3.2 *Estrazione del DNA*

La macinazione del materiale vegetale è stata effettuata tramite *TissueLyser* Retsch® (QIAGEN, Haan, Germania) (Figura 8). Questo strumento permette la distruzione simultanea di molti campioni biologici scuotendoli velocemente e vigorosamente con sferette metalliche atte a tritare il tessuto vegetale, condizione necessaria all'estrazione di DNA. A ciascuna provetta contenente i tessuti verdi è stata aggiunta una pallina di tungsteno. Le provette così preparate, in gruppi di 24, sono state inserite negli appositi adattatori (per un totale di 48 campioni). Quest'ultimi sono stati collocati a -80°C per circa 1 ora in modo da congelare il tutto ed evitare il surriscaldamento del materiale durante la macinazione. Dopo questa fase, la macinazione è avvenuta per 30 secondi alla

frequenza di 28 Hz. Qualora alcuni dei campioni non fossero completamente triturati, la macinazione è stata estesa di 30 secondi e comunque fino a completa polverizzazione dei tessuti. Il materiale così prodotto è stato stoccato a -80°C fino al momento dell'estrazione del DNA.



Figura 8. TissueLyser, strumento che permette la macinazione simultanea di più campioni.

L'estrazione del DNA è stata condotta sull'intera collezione di 428 campioni utilizzando il kit *GeneElute Plant Genomic DNA Miniprep* (Sigma-Aldrich, St Louis, MO), seguendo il protocollo fornito. La procedura di estrazione si articola in 9 passaggi (Figura 9), in breve:

1. lisi cellulare, al fine di favorire la dissoluzione delle membrane e la liberazione del DNA;
2. precipitazione, che favorisce la formazione di un *pellet* contenente i vari detriti, proteine e polisaccaridi;

3. filtraggio, ovvero il recupero del surnatante formatosi in seguito ad una centrifuga alla massima velocità e passaggio in una colonnina filtrante in modo da eliminare ogni detrito rimasto;
4. ligazione del DNA alla membrana di silice fornita nel kit, favorendone l'affinità di legame con soluzioni apposite;
5. lavaggio del DNA tramite una soluzione alcolica allo scopo di rimuovere i contaminanti e trattenere il solo DNA genomico;
6. eluzione del DNA. Allo scopo di massimizzare la raccolta di DNA pur mantenendo una concentrazione quanto più elevata possibile, sono state effettuate due eluizioni successive e separate in 60 μ l e 40 μ l.

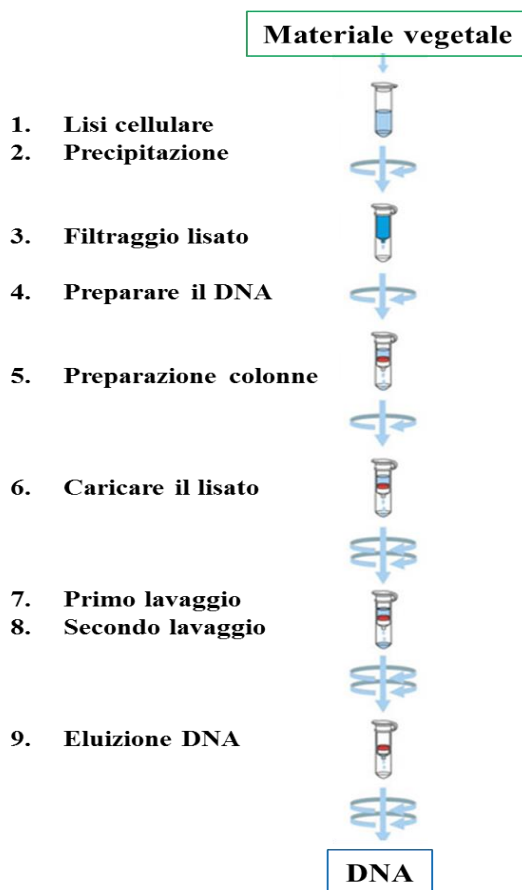


Figura 9. Immagine che illustra i vari passaggi del protocollo di estrazione del DNA (GeneElute Plant Genomic DNA Miniprep) tramite colonnine.

3.3.3 Controllo qualità e quantità del DNA

La qualità dell'estrazione e la concentrazione di DNA nei campioni è stata valutata tramite corse elettroforetiche. Il *gel* d'agarosio all'1% è stato preparato utilizzando come *buffer* il TBE. Ad ogni corsa, insieme ai campioni (12 μ l volume totale, di cui 2 μ l di DNA e 10 μ l di *loading dye*), è stato caricato DNA di Fago λ a diverse concentrazioni note (5, 10, 20 ng/ μ l), in modo da poter stimare la concentrazione di DNA presente nei vari campioni. La qualità dell'estrazione è stata valutata osservando la compattezza delle bande. Il *gel* doveva dunque mostrare bande non frammentate, cioè un'unica banda

all'inizio di ogni pozzetto che stava a indicare l'integrità del DNA genomico. Le corse elettroforetiche sono state effettuate a 60 V.

L'uso di uno spettrofotometro è stato affiancato alla corsa elettroforetica per rendere più accurata la quantificazione del DNA estratto. Ciascun campione è stato letto nel volume di 2 μ l, annotando la concentrazione stimata a partire dall'equazione di Beer-Lambert su un cammino ottico di 1 mm. Inoltre, considerando che il DNA ha il picco di assorbimento a 260 nm, la qualità di ogni campione è stata valutata secondo due rapporti: a) 260/280, che indica la purezza della preparazione degli acidi nucleici, perché le proteine assorbono a 280 nm e b) 260/230 che riflette la contaminazione da parte di carboidrati, fenoli, peptidi. Per ottenere campioni di buona qualità il primo rapporto dovrebbe essere circa 1,8 mentre il secondo dovrebbe avere valori intorno a 2.

3.3.4 Normalizzazione dei campioni

Il protocollo di sequenziamento richiedeva almeno 600 ng/ μ l di DNA. Numerosi campioni, però, risultavano non essere sufficientemente concentrati. Di conseguenza, è stato recuperato dal *backup* nuovo tessuto vegetale il quale è stato nuovamente sottoposto al protocollo di estrazione descritto nel paragrafo 3.3.2. Se anche dopo l'ulteriore estrazione alcuni campioni continuavano a non essere idonei, questi sono stati sottoposti ad una procedura di concentrazione. Dopo aver riunito le due eluizioni (60 e 40 μ l) e la nuova estrazione del campione, le provette sono state inserite nell'evaporatore Concentrator 5301® (Eppendorf AG, Amburgo, Germania). Questa macchina permette l'evaporazione della soluzione acquosa e quindi la formazione del *pellet* di DNA. L'acido nucleico è stato nuovamente eluito in un volume minore, di modo da ottenere la concentrazione richiesta dal protocollo di sequenziamento.

I campioni così concentrati sono stati nuovamente quantificati, ed assieme agli altri sono stati normalizzati alla concentrazione di 20 ng/μl in un volume di almeno 50 μl su piastre da 96 pozzetti e inviati a IGA Technologies Services (Udine, Italia) per la produzione di dati molecolari. Prima di essere spedite, su ciascuna piastra sono stati inseriti dei controlli replicati.

3.4 Genotipizzazione RADSeq

La procedura di RADSeq non era mai stata applicata su *T. urartu*. Pertanto, si è reso necessario testare e mettere a punto il protocollo su un *subset* di 24 campioni prima di estenderlo all'intera collezione. Non essendo presente alcuna informazione pregressa sulla variazione molecolare presente all'interno della collezione, il *subset* di 24 campioni è stato composto da 13 accessioni geograficamente distanti ed 11 accessioni raccolte in reciproca prossimità. Tra queste ultime, è stata inserita l'accessione sequenziata di *T. urartu* (G1812; nella collezione corrisponde a SSA_324). La localizzazione geografica è stata utilizzata come predittore della diversità genetica. E' infatti possibile supporre che la localizzazione geografica delle popolazioni naturali influenzi la loro differenziazione. Accessioni più vicine saranno più simili, accessioni più distanti saranno più diverse. Anche se non sempre questi presupposti sono mantenuti a causa di campionamento genetico, migrazione, selezione (Dell'Acqua *et al.* 2014) questo presupposto è un punto di partenza ragionevole in una collezione di materiale inesplorato. Siccome questo primo esperimento era soprattutto volto a mettere a punto la tecnica di sequenziamento, questo mirava ad ottenere una visione approssimativa del numero minimo e massimo di polimorfismi presenti all'interno della collezione. Una volta testato il protocollo su questo piccolo *subset*, la procedura di sequenziamento è stata estesa a tutta la collezione (in corso al momento della tesi).

Il protocollo di RADSeq (Figura 10) prevede l'utilizzo di due enzimi di restrizione, l'uno a bassa frequenza di taglio, *SphI-HF* (sito di taglio a 6 nucleotidi, 5'-GCATG[^]C-3') e l'altro ad alta frequenza di taglio, *MboI* (sito di taglio a 4 nucleotidi, 5'-[^]GATC-3'). Il DNA genomico di ciascun campione è esposto ai due enzimi di restrizione in una reazione indipendente. Ai frammenti di restrizione originati dal taglio enzimatico sono attaccati due adattatori complementari alle sequenze presenti sulla *flowcell* della piattaforma di sequenziamento. L'adattatore P1 presenta un *barcode*, ovvero una sequenza di nucleotidi che permette di assegnare in maniera univoca una sequenza a un determinato individuo, il secondo adattatore (P2) presenta una struttura a Y divergente che non gli consente di legarsi al *primer* P2 a meno che non sia terminata l'amplificazione dell'adattatore P1. Ciò garantisce che tutti frammenti amplificati presentino l'adattatore P1 con il *barcode*, il sito di restrizione parziale, poche centinaia di basi della sequenza fiancheggiante e l'adattatore P2. L'adesione tra gli adattatori e le sequenze *sticky-end* (ovvero sequenze con estremità adesive che permettono una più semplice ligazione degli adattatori), è effettuata in un termociclatore a 94 °C per circa 2 min, dopo di che la temperatura è abbassata di 1° al minuto fino a raggiungere i 20 °C (in modo da consentire l'adesione degli adattatori a tutti i frammenti di DNA precedentemente denaturati). È stata prelevata un aliquota da ogni campione, purificata dagli eventuali adattatori che non hanno reagito e infine riunito il tutto in una provetta. La presenza del *barcode* rende possibile l'unione di tutti i campioni rendendo la procedura più veloce ed efficace. Una prima selezione, basata sulla dimensione dei frammenti, è stata effettuata tramite corsa elettroforetica. Il taglio del gel a determinate dimensioni assicura di ottenere sequenze dello stesso peso molecolare in tutti i campioni, e quindi avere, dopo il sequenziamento, *reads* rappresentative di quella parte di genoma.

In questo caso sono stati effettuati due tagli:

- 1) Il primo tra 300-450 pb;
- 2) Il secondo tra 450-650 pb.

Le sequenze così ottenute sono state eluite e sottoposte a PCR, i *primer* complementari ai due adattatori andranno ad amplificare solo le sequenze che presentano entrambi i tipi; in questa maniera viene effettuata una seconda selezione.

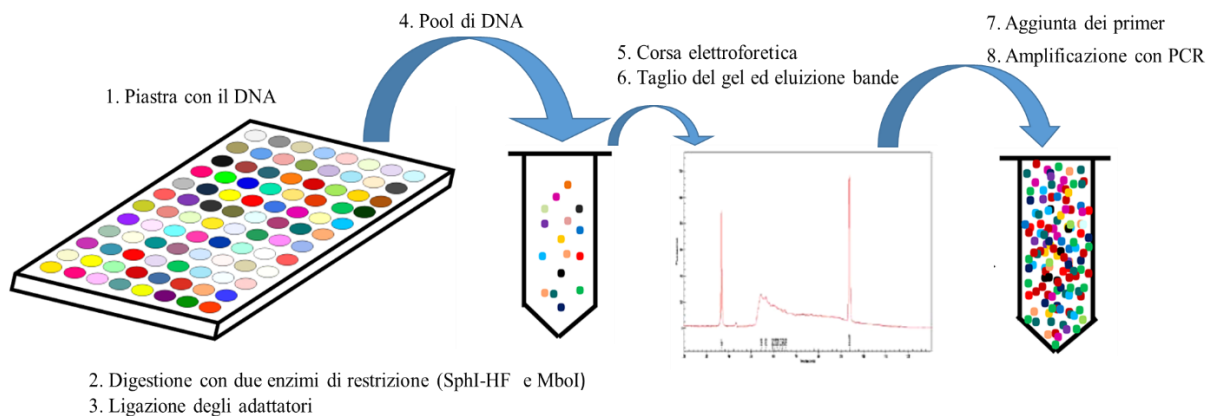


Figura 10. Passaggi del protocollo di RADSeq: 1) Piastra contenente DNA, enzimi di restrizione e adattatori; 2-3) Digestione dei campioni con i due enzimi di restrizione SphI-HF e MboI e ligazione degli adattatori ai frammenti di DNA *sticky-end*; 4) Pulizia dei campioni in modo da eliminare gli adattatori che non hanno reagito; 5-6) Corsa elettroforetica, taglio ed eluizione di specifiche bande del gel; 7) Aggiunta dei *primer* complementari agli adattatori; 8) Amplificazione con PCR. Modificata da Elshire *et al.* 2011.

La libreria ottenuta è stata sottoposta a sequenziamento con la piattaforma Illumina HiSeq™ 2000 (Illumina®, Solexa, San Diego, CA, USA). Dopo la costruzione della libreria, questo metodo di *next generation sequencing* si articola in due passaggi (Figura 11):

1. *Bridge* PCR, cioè l'amplificazione a ponte dei frammenti di DNA attaccati alla *flowcell* per creare dei *cluster* contenenti un numero elevato di copie dello stesso frammento da sequenziare. Questo è un passaggio necessario che

all'incorporazione dei nucleotidi permette di creare un segnale luminoso abbastanza forte da essere catturato dalla camera;

2. Sequenziamento, basato sull'incorporazione di 4 nucleotidi marcati rispettivamente con 4 fluorofori in grado di interrompere temporaneamente la polimerizzazione e la DNA polimerasi. Incorporato il nucleotide avviene la lettura, successivamente viene eliminato il fluoroforo ed il gruppo terminatore al 3' (che impedisce l'attacco di un altro nucleotide) e il ciclo si ripete.

Il risultato finale è la produzione di milioni di *reads*.

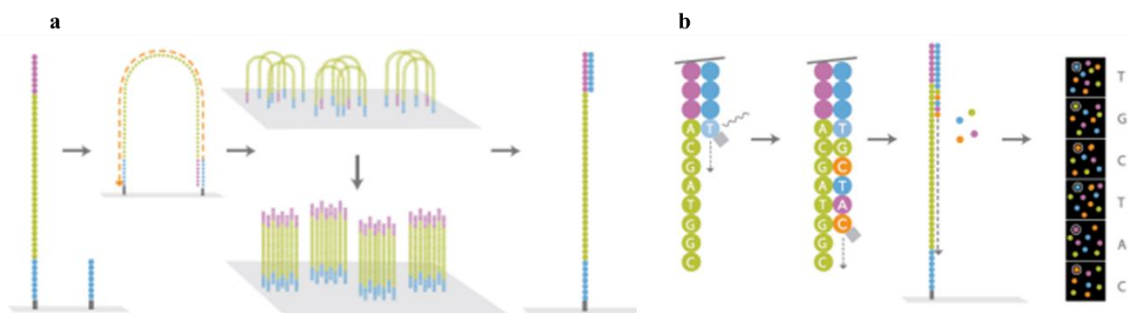


Figura 11. Passaggi del sequenziamento con piattaforma Illumina: **a)** *Bridge* PCR, amplificazione in cluster del frammento di DNA da sequenziare in modo da ottenere un segnale abbastanza forte da essere rilevato con la camera; **b)** Sequenziamento, con 4 nucleotidi terminatori marcati con 4 fluorofori differenti, modificato da <http://www.medizinische-genetik.de/>.

Una volta ottenute le sequenze, i *barcode* vengono utilizzati per fare risalire ad ogni campione le corrispondenti *reads*. Queste vengono poi sottoposte a *trimming*, cioè vengono eliminati gli adattatori in modo da consentire il mappaggio di quest'ultime contro il genoma di riferimento. Le sequenze allineate vengono utilizzate per chiamare i polimorfismi in forma di SNP sulla base della diversità nucleotidica dei diversi campioni in una data posizione. Il programma Stacks (Catchen *et al.* 2013) è stato utilizzato per allineare le *reads* prodotte dai 24 campioni contro il genoma di riferimento G1812. Per questa prima analisi, che aveva lo scopo di testare bontà e solidità del protocollo, sono

stati usati filtri di qualità molto rigidi. Il *coverage* minimo per chiamare un polimorfismo è stato impostato a 2, ed il tasso di fallimento massimo per considerare un marcatore valido è stato impostato al 30%. Inoltre, vista l'elevata componente di DNA ripetuto attesa nel genoma di *T. urartu*, e vista la modalità di riproduzione autogama di questa specie, tutti i marcatori eterozigoti sono stati eliminati come possibile risultato di allineamenti imperfetti. Infine, in questa prima fase, sono stati considerati solo i loci con un solo nucleotide discriminatorio: la presenza di più SNP in serie in posizioni molto ravvicinate potrebbe altresì sottendere ad un allineamento multiplo delle *reads*. Tutti questi parametri, riconosciuti come prudenziali all'eccesso, saranno valutati e rilassati opportunamente nell'analisi dell'intera collezione. Tali criteri stringenti, producono il *set* di marcatori più solidi a disposizione per valutare la variabilità presente nel campione iniziale.

3.4.1 Analisi dati molecolari

I dati prodotti dal sequenziamento, ottenuti come file di testo, sono stati formattati con R in modo da risultare leggibili dal programma Tassel 4.0 (Bradbury *et al.* 2007). Con questo software è stato possibile effettuare un'analisi di diversità sul *subset* di 24 campioni utilizzati per testare il protocollo di sequenziamento. La diversità molecolare tra i campioni analizzati è stata prima di tutto descritta da un'analisi filogenetica di *neighbor joining*. La ricerca del *neighbor* cioè del “vicino” di ciascuna specie sull'albero filogenetico viene effettuata attraverso un processo di minimizzazione della somma di tutte le distanze fra le specie considerando tutte le coppie possibili. In breve, si costruisce una matrice di diversità a partire dai dati molecolari e sulla base di questa matrice si valuta quale siano i due campioni che presentano la relazione più forte rispetto a tutti gli altri. A partire poi da questa coppia vengono uniti in maniera progressiva gli altri campioni,

rappresentando graficamente un albero filogenetico. Per fornire una visualizzazione sintetica del dato molecolare in funzione della provenienza dei campioni, sul *subset* di SNP sono state estratte le componenti principali (PC) tramite una *principal component analysis* (PCA) tramite il programma Tassel 4.0 (Bradbury *et al.* 2007). Le coordinate della prima e seconda PC sono state caricate in R ed utilizzate per produrre un *output* grafico.

3.5 Campo sperimentale

3.5.1 La semina

Il campo (20m x 40m) è stato seminato l'11 Dicembre 2014 al Centro di Ricerca Centro Interuniversitario sulle Biomasse da Energia (CRIBE) in San Piero a Grado (Pisa, Italia, 43°40'16.37"N / 10°20'52.47"E). Dall'intera collezione così come sono state ricevute dalle *seedbank* sono state selezionate e conservate 5 spighette come *backup*, mentre il restante materiale è stato seminato. Dipendentemente dall'origine del germoplasma, la quantità di seminato è stata variabile, dalle 10 alle 30 spighette: questo è dovuto al fatto che tra gli scopi del campo vi era la produzione di quanto più seme possibile per consolidare la collezione. Le spighette sono state seminate manualmente su 20 righe e 22 colonne, subito dopo il suolo è stato rullato. Oltre all'intera collezione di *T.urartu* sono stati inclusi anche tre *T. boeoticum* e tre *T. monococcum*, utilizzati come controllo diploide. Inoltre, per evitare l'effetto bordo, il campo è stato circondato da due file di *T. aestivum*. Una concimazione di 25 kg/ha di N è stata effettuata subito dopo la semina e replicata in fase di levata. Il diserbo è stato condotto a mano.

3.5.2 *Analisi fenotipiche*

Allo scopo di osservare la variabilità della collezione sono stati raccolti 8 caratteri fenotipici ai corrispondenti stadi fenologici:

- portamento, rilevato a inizio accostamento valutando su una scala che va da 1 a 5 la crescita delle parti aeree della pianta. Il valore più basso indica un portamento prostrato e quello più alto un portamento eretto;
- epoca fioritura, segnando la data di fioritura quando più del 50% degli individui di un'accessione erano spigati;
- altezza della pianta a spigatura, misurando 3 individui per filetta dalla base della pianta fino alle spighe, escludendo le reste;
- area foglia bandiera a spigatura, misurando 3 individui per accessione. Con un righello è stata misurata la lunghezza e la larghezza della foglia bandiera (l'ultima foglia del culmo che racchiude l'infiorescenza);
- lunghezza della spiga, selezionando 3 individui per accessione e misurando la lunghezza della spiga escludendo le reste;
- numero spighette per spiga, selezionando 3 individui per accessione e contando manualmente il numero delle spighette per spiga.

È stata inoltre annotata l'assenza o la presenza di reste e la pelosità della spiga.

3.5.3 *Gestione dati fenotipici*

I dati fenotipici sono stati esportati come file testuali ed importati in R (R Development Core Team 2015). Le analisi di distribuzione sono state effettuate su sei degli otto caratteri fenotipici, escludendo i dati riguardanti la presenza o assenza di reste e la pelosità della spiga che mostravano insufficiente variabilità. La distribuzione dei campioni secondo i

caratteri fenotipi rilevati è stata rappresentata tramite istogrammi con la funzione *hist()*. Con il pacchetto R/corrplot (Wei 2013) è stato calcolato il grado di correlazione dei dati fenotipici all'interno della collezione, in modo da valutare quali e soprattutto in quale misura i caratteri risultano tra loro correlati. Sempre su questi caratteri è stata condotta una PCA con la funzione *princomp()* di R. Analogamente a quanto fatto con i dati climatici, la PCA è stata utilizzata per osservare la distribuzione spaziale della variabilità fenotipica ed identificare eventuali raggruppamenti all'interno della collezione.

4 Risultati

4.1 Geografici

L'analisi geografica GIS è stata utilizzata per evidenziare le differenze ambientali dei siti di campionamento delle accessioni di *Triticum urartu*. Il software QGIS è stato utilizzato con il dataset *BioClim* per creare cartine geografiche della Mezzaluna Fertile che forniscano una visione d'insieme della distribuzione delle accessioni secondo altitudine (Figura 12), temperatura (Figura 13) e piovosità annuale (Figura 14). Queste tre variabili sono state scelte in quanto le più rappresentative del cline ecologico di interesse. Le analisi geografiche hanno evidenziato che la maggior parte delle accessioni cresce in regioni pianeggianti con altitudini non superiori ai 500 m (colore verde nella mappa in Figura 12), tuttavia alcuni campioni crescono ad altitudine più elevate (colore rosso nella mappa in Figura 12), fino ad arrivare a un'altezza massima di 2285 m.

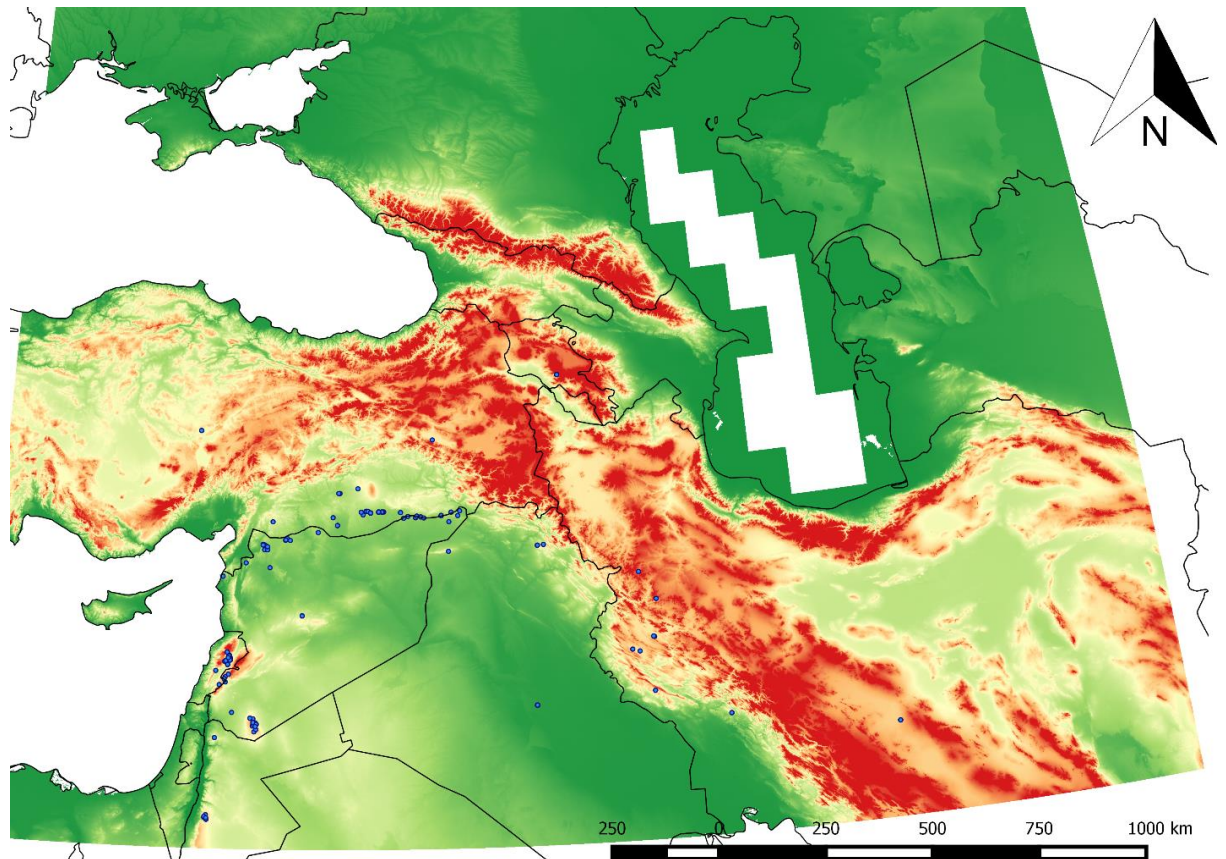


Figura 12. Mappa che raffigura l'altitudine della Mezzaluna Fertile, le accessioni della collezione sono rappresentate dai cerchi blu. La variazione di colore indica il *range* di altitudine in cui si trovano i campioni, da verde (-28 a 550 m), a giallo (1128 m) e sino al rosso, con un massimo di 2285 metri.

In accordo con il gradiente altitudinale, troviamo che la maggioranza dei campioni della collezione vive in un *range* di temperatura elevato (Figura 13). Tuttavia, le temperature medie annuali presentano una variazione molto ampia, dai 4 ai 23° C, con una media di 16° C e deviazione standard di 1,3. Questo è in accordo con la marcata variazione altitudinale dei siti di campionamento. Per quanto riguarda la piovosità annuale (Figura 14), si riscontra una distribuzione della collezione concentrata in regioni con piovosità bassa, inferiore ai 500 mm l'anno (zone chiare della mappa in Figura 14). Tuttavia, esistono accessioni che crescono in zone con più di 1000 mm di pioggia l'anno, ovvero regioni con elevata piovosità. L'insieme di questi dati geografici macroscopici

mostra come la variazione climatica nell'area di campionamento è molto vasta, e conferma che la collezione di *T. urartu* comprende popolazioni che crescono in zone climatiche diverse tra loro.

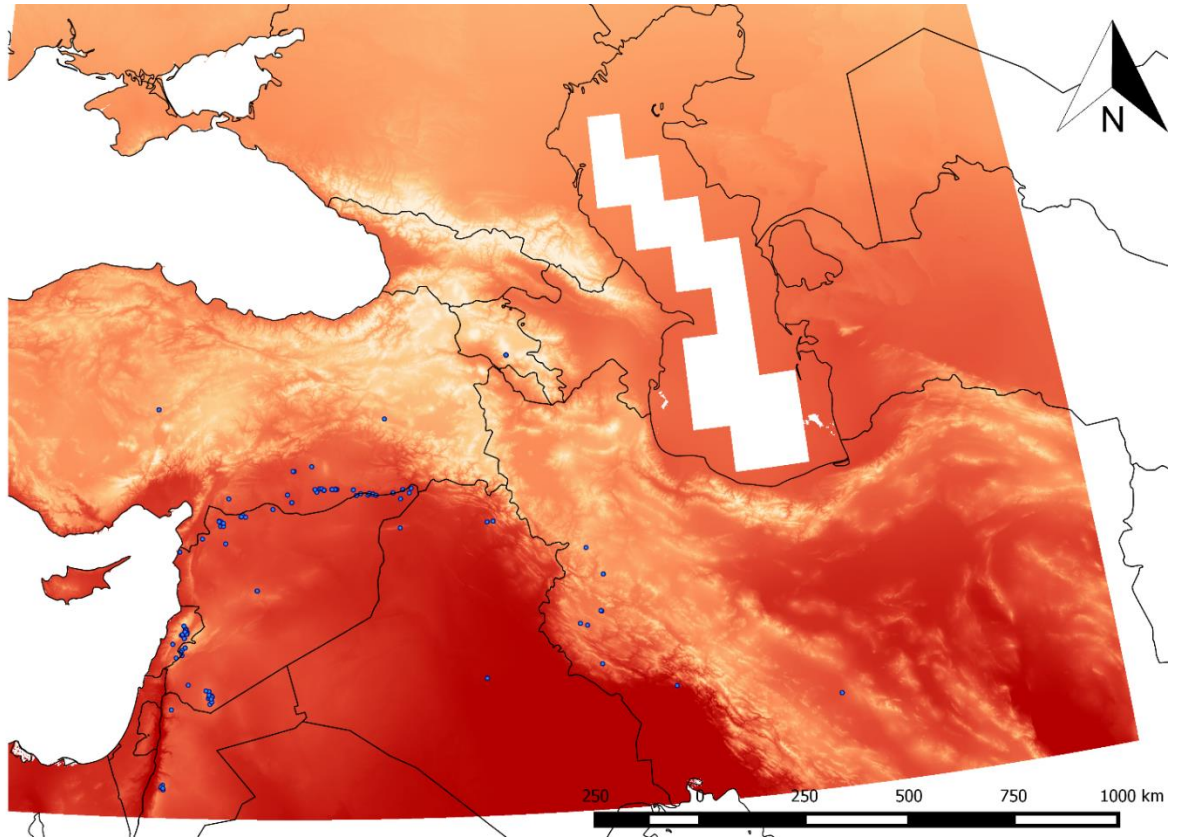


Figura 13. Mappa che raffigura la temperatura media annuale degli anni 1950-2000 della Mezzaluna Fertile, le accessioni della collezione sono rappresentate dai cerchi blu. La variazione di colore indica il range di temperatura in cui si trovano i campioni, da bianco a rosso per valori più elevati.

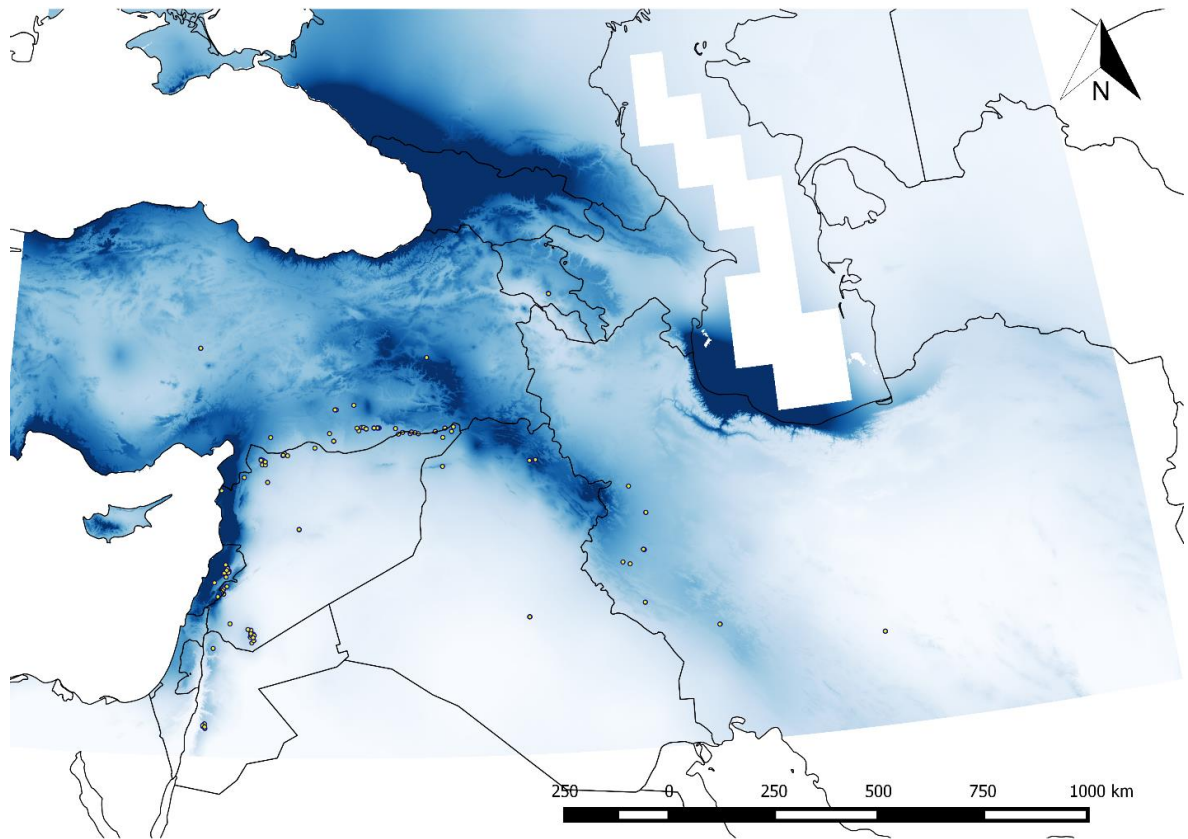


Figura 14. Mappa che raffigura la piovosità annuale media degli anni 1950-2000 della Mezzaluna Fertile, le accessioni della collezione sono rappresentate dai cerchi gialli. La variazione di colore indica il range di piovosità in cui si trovano i campioni, da bianco a blu per regimi di pioggia crescenti.

Le condizioni ecologiche di ciascun sito di campionamento sono state ottenute tramite DIVA-GIS. Una volta associate le variabili *Bioclim* a ciascun campione, la funzione *boxplot* di R è stata usata per ottenere una rappresentazione grafica della distribuzione delle accessioni secondo le 19 variabili bioclimatiche. In Figura 15 sono rappresentate le distribuzioni per le variabili bioclimatiche legate alla temperatura. La variabile BIO₄ (stagionalità delle temperature, misurata come deviazione standard delle temperature annuali), è stata modificata dividendola per 100 di modo da renderla graficabile con le altre variabili. I valori di BIO₁, ovvero la temperatura media annuale, avvalorano i risultati di Figura 13, ovvero che i campioni crescono in un *range* di temperatura annuale elevato. Il *boxplot* infatti evidenzia che il terzo e quarto quartile della

distribuzione sono raccolti tra i 18° C ed i 22° C. Primo e secondo quartile della distribuzione hanno un *range* più elevato, che si spinge sino a meno di 10° C.

BIO₆, la temperatura minima del mese più freddo, indica che la maggior parte dei campioni è distribuita tra valori di temperatura vicini allo zero. Anche in questo caso la distribuzione non è normale, ed il terzo quartile risulta essere notevolmente più ampio rispetto al secondo. Di notevole interesse è inoltre la presenza di numerosi *outlier* sotto il primo quartile, che sta a indicare che numerose accessioni riescono a crescere in regioni con inverni rigidi (con massimo valore minimo inferiore a -10° C). Questo dato è confermato anche dalla distribuzione della variabile BIO₁₁, che rappresenta la media della temperatura del quartile più freddo. Anche per questa variabile è possibile osservare la presenza di numerose accessioni che crescono in zone con inverni molto rigidi. La variabilità della collezione in rapporto ai diversi aspetti della temperatura è quindi molto elevata.

Misure di Temperatura

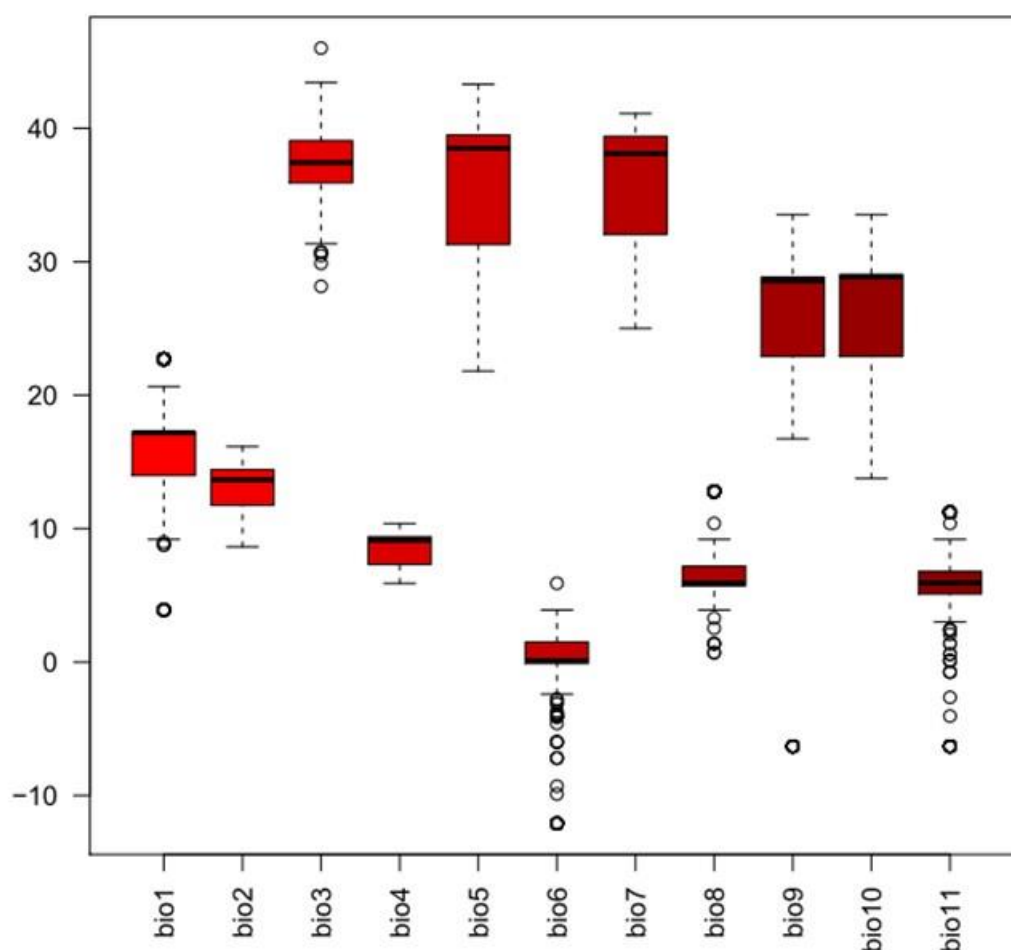


Figura 15. *Boxplot* che rappresenta la distribuzione dei campioni raccolto nella Mezzaluna Fertile secondo la temperatura. Sull'ascissa sono riportate le variabili bioclimatiche della temperatura (BIO₁ = Temperatura media annuale, BIO₂ = Media della variazione giornaliera [media del mese T°max -T°min], BIO₃ = Isotermalità [BIO₂/BIO₇] [*100], BIO₄ = Temperatura stagionale dato modificato [deviazione standard / 100], BIO₅ = Temperatura max del mese più caldo, BIO₆ = Temperatura min del mese più freddo, BIO₇ = Variazione delle Temperatura annuale [BIO₅ - BIO₆], BIO₈ = Media della Temperatura del quartile più umido, BIO₉ = Media della Temperatura del quartile più secco, BIO₁₀ = Media della Temperatura del quartile più caldo, BIO₁₁ = Media della Temperatura del quartile più freddo), sull'ordinata è riportata la temperatura in C°.

Le variabili bioclimatiche legate alla piovosità aiutano a completare il quadro della variabilità climatica inclusa nella collezione (Figura 16). Osservando BIO₁₂, ovvero la precipitazione media annuale rappresentata anche in Figura 14, si evince che la

maggioranza delle accessioni è distribuita in regioni con piovosità inferiore ai 500 mm. Ciò nonostante, alcune accessioni crescono a valori estremi di piovosità, rispettivamente nella parte inferiore e superiore della distribuzione. La distribuzione dei valori medi di piovosità annuale è notevolmente ampia, suggerendo che queste accessioni sono dei buoni candidati per l'adattamento all'uno o all'altro estremo delle condizioni di piovosità. I *boxplot* di BIO₁₄, BIO₁₇ e BIO₁₈, cioè le precipitazioni del mese e del quartile più secco e del quartile più caldo, sono costituiti da valori di piovosità estremamente bassi rispetto alle altre variabili e per questo motivo la loro variazione non è apprezzabile da questo grafico.

Misure di Piovosità

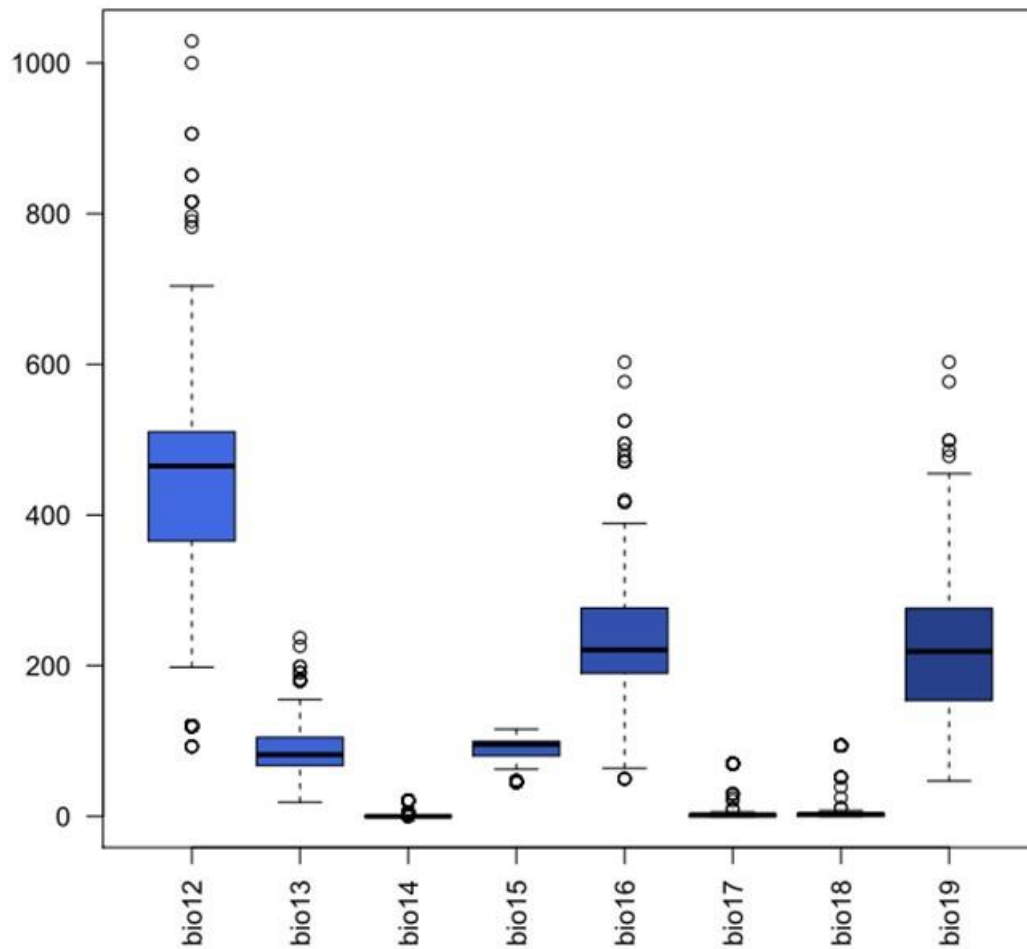


Figura 16. *Boxplot* che rappresenta la distribuzione dei campioni raccolto nella Mezzaluna Fertile secondo la piovosità. Sull'ascissa sono riportate le variabili bioclimatiche della piovosità (BIO₁₂ = Precipitazione annuale BIO₁₃ = Precipitazione del mese più umido, BIO₁₄ = Precipitazione del mese più secco, BIO₁₅ = Precipitazione stagionale, BIO₁₆ = Precipitazione del quartile più umido, BIO₁₇ = Precipitazione del quartile più secco, BIO₁₈ = Precipitazione del quartile più caldo, BIO₁₉ = Precipitazione del quartile più freddo), sull'ordinata sono riportati i valori della piovosità in mm di pioggia.

I valori di piovosità e temperatura sono strettamente legati al gradiente altitudinale in cui le accessioni sono state campionate. La gran parte dei campioni si trova in regioni ad altitudine inferiore ai 500 m (Figura 17), molti dei quali al livello del mare. La collezione di *T. urartu*, comunque, raccoglie numerose accessioni campionate ad

altitudini tra i 1000 ed i 1500 metri, e sino a più di 2000 metri. Ai due capi della distribuzione si contrappongono estati calde e siccitose ed inverni freddi e piovosi, come riportato in Figura 16.

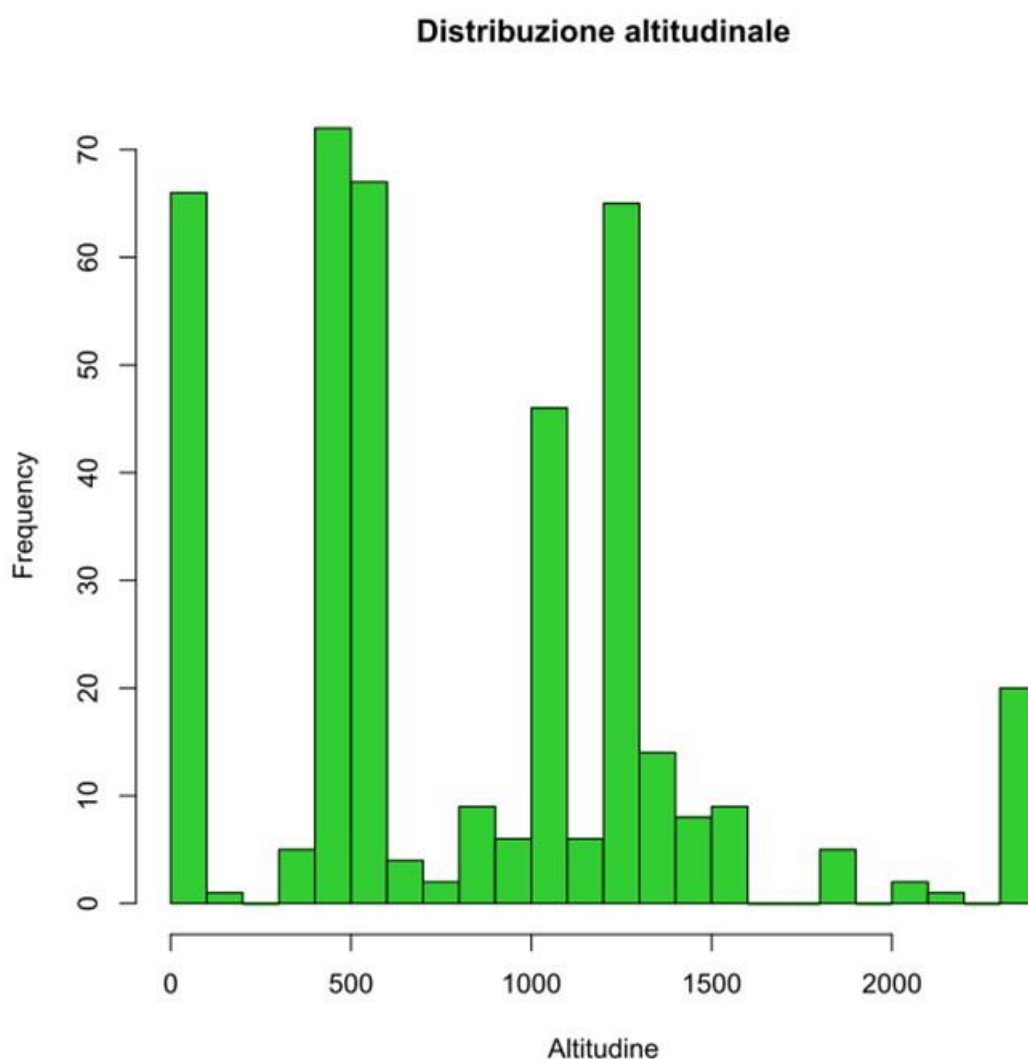


Figura 17. Istogramma che rappresenta la distribuzione dei campioni raccolti nella Mezzaluna Fertile secondo l'altitudine. Sull'ascissa è riportato il *range* di altitudine a cui sono state campionate le accessioni, in ordinata è riportata la frequenza di distribuzione dei campioni.

La correlazione tra i dati bioclimatici tra loro e con l'altitudine, comunque, non sempre è lineare come ci si potrebbe aspettare da fenomeni intercorrelati quali quelli climatici (Figura 18). Come prevedibile, l'altitudine è correlata negativamente con quasi

tutte le misure di temperatura (BIO₁, BIO₂, BIO₅, BIO₆, BIO₉-BIO₁₁) mentre è interessante notare la scarsa correlazione esistente con le variabili BIO₄, BIO₇ e BIO₈ (Temperatura stagionale, variazione della temperatura annuale e media della temperatura del quartile più umido, rispettivamente) e la mancanza di correlazione con BIO₃ (Isotermalità).

Dal grafico delle correlazioni è possibile notare inoltre una netta divisione della figura. Il primo gruppo ad emergere è costituito dalle variabili di temperatura contro se stesse (BIO₁-BIO₁₁), il secondo da quelle di piovosità (BIO₁₂-BIO₁₈) contro la temperatura e l'ultimo gruppo dalle variabili di piovosità contro se stesse. Nella maggior parte dei casi quando gruppi di variabili affini sono confrontate contro se stesse risultano essere positivamente correlate, mentre quando vengono messe a confronto le variabili di temperatura con quelle di piovosità si osservano correlazioni negative. Come per l'altitudine sono presenti delle variabili che non rispecchiano i risultati attesi ma che sono in netto contrasto, osservando BIO₃, infatti, è possibile notare come questa variabile risulti essere poco correlata con tutte le altre variabili di temperatura e addirittura negativamente correlata con BIO₄ e in misura minore con BIO₇.

Un'altra variabile che mostra risultati discordanti è BIO₁₅, questa infatti risulta correlata positivamente con alcune variabili di temperatura (BIO₃, BIO₆, BIO₉ e BIO₁₁) e negativamente con quelle di piovosità. Queste variabili rappresentano la temperatura dei mesi e quartili più freddi e umidi, la correlazione positiva della precipitazione stagionale con queste variabili indica un aumento delle piogge nei periodi in questione che porta, di conseguenza, un abbassamento delle temperature. Il disegno generale che emerge da questa analisi è quello di un cline ecologico complesso, che attraversa diverse regioni

climatiche che presentano combinazioni alternative delle misure di altitudine, temperatura e piovosità.

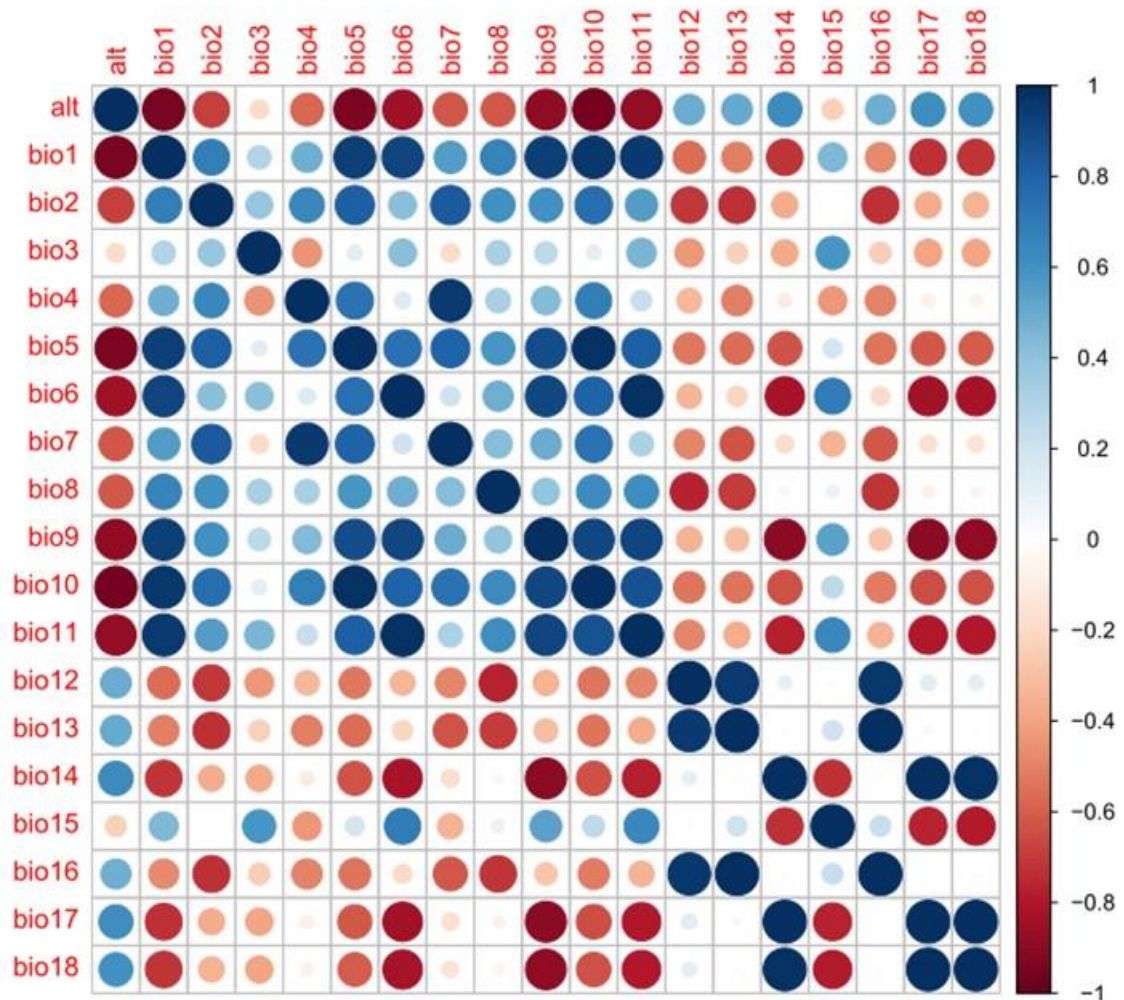


Figura 18. Immagine che rappresenta la correlazione tra le variabili bioclimatiche dei campioni, sulla destra si trova una scala che indica il livello di correlazione delle variabili. La dimensione dei cerchi indica il grado di correlazione, grandi dimensioni forte correlazione; il colore dei cerchi indica in che modo sono correlate due variabili, blu positivamente e rosso negativamente. Alt = Altitudine per le variabili bioclimatiche vedi Tabella 3.

In questo quadro complesso, un'analisi di PCA può essere utile a condensare la variazione climatica dell'area di raccolta in poche variabili ad elevata informatività

(Figura 19). La PCA sui dati climatici restituisce le prime due PC responsabili rispettivamente dell'86% e dell'11% della varianza del *dataset*. Il primo risultato che si evince dall'analisi è che non è presente una netta divisione dei campioni secondo i paesi di campionamento, ma che è altresì presente un'elevata variabilità ambientale non solo all'interno della collezione ma anche tra le varie accessioni di una determinata regione. Dall'immagine è possibile apprezzare la presenza di un primo gruppo, distribuito su valori di PC1 e PC2 negativi che comprende principalmente le accessioni siriane, iraniane e giordane. In questa area dello spazio PC si trovano anche i campioni armeni, sebbene siano raccolti in valori estremamente negativi di PC1.

Considerando la distribuzione secondo l'asse di PC2 troviamo a valori positivi due raggruppamenti. Il primo si trova a valori di PC1 negativi e comprende la maggior parte di accessioni provenienti dal Libano e alcune provenienti dalla Siria e dall'Iraq; il secondo gruppo si trova a valori positivi di PC1 ed è costituito da altre accessioni siriane, la maggioranza di quelle turche e alcune irachene.

E' interessante notare come sia difficile identificare dei *cluster* evidenti nello spazio di PC; piuttosto, si può vedere come siti di campionamento contigui presentino condizioni bioclimatiche molto diverse. Viceversa, siti di campionamento lontani (come ad esempio Iran e Giordania, o Turchia e Libano) presentano condizioni bioclimatiche coincidenti nella prima e seconda PC, che insieme condensano il 97% della varianza del *dataset* originale. Questo risultato conferma che la variazione climatica non è direttamente correlata con la variazione spaziale: in altre parole, a maggiore distanza geografica non corrisponde maggiore distanza ecologica.

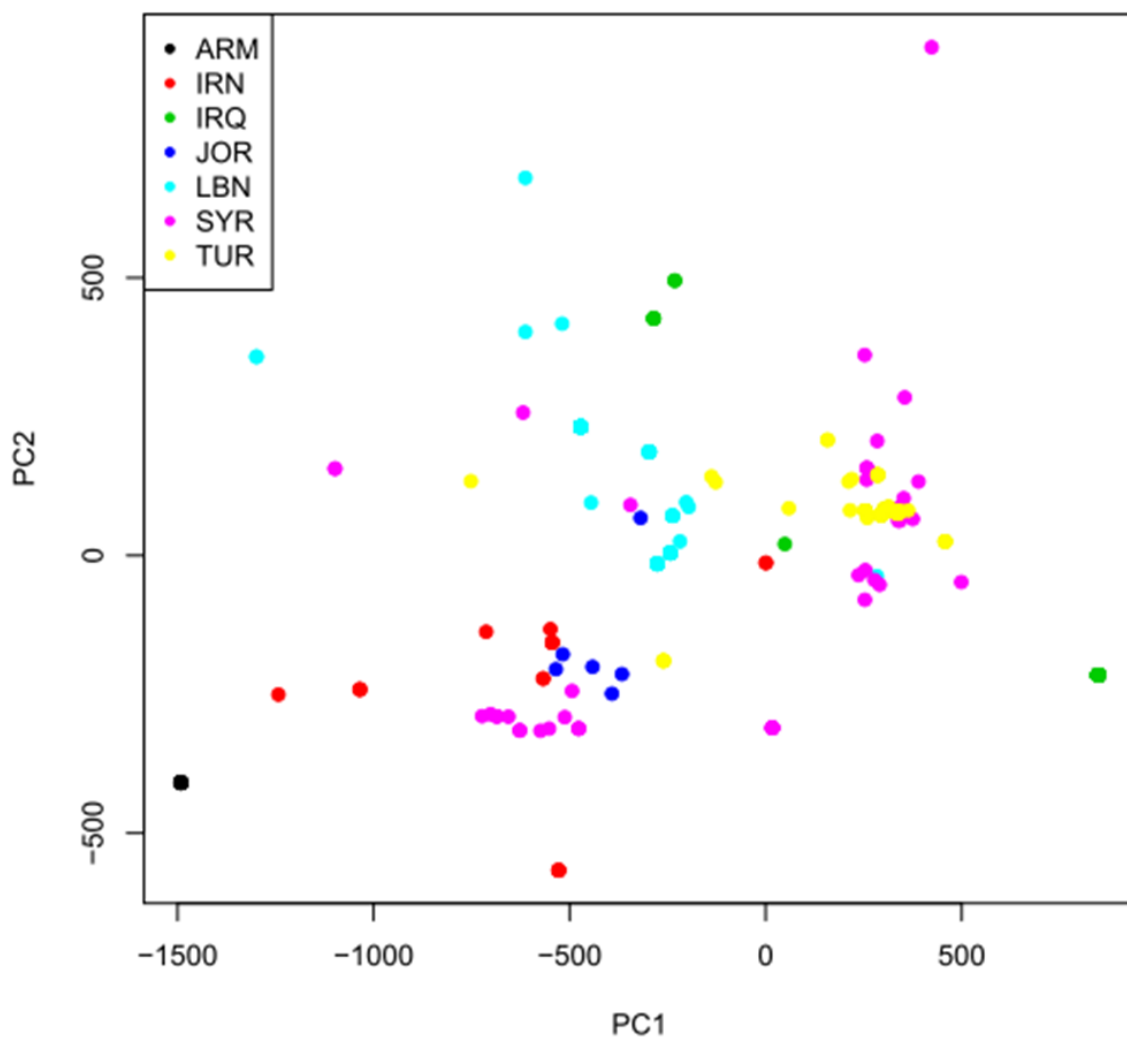


Figura 19. PCA delle variabili bioclimatiche secondo le nazioni di campionamento delle accessioni. L'immagine rappresenta le due componenti principali che spiegano rispettivamente il 86% e l'11% della varianza del *dataset* bioclimatico. La legenda in alto a sinistra indica la provenienza dei campioni, a ogni regione appartiene un colore diverso (ARM= Armenia, IRN = Iran, IRQ = Iraq, JOR = Giordania, LBN = Libano, SYR = Siria, TUR = Turchia).

Allo scopo di comprendere quali fossero le variabili originali ad avere il maggior peso nella determinazione dei primi tre assi di PC, è stata compiuta un'analisi di correlazione tra loro (Figura 20). La PC1, che spiega l'86% della varianza del *dataset*, è contribuita dalla maggior parte delle variabili correlate tra loro in Figura 18. I valori

altitudinali mostrano una forte correlazione inversa con la prima PC, cioè all'aumentare dei valori della PC1 diminuisce l'altitudine a cui le accessioni sono state campionate. Prevedibilmente, le variabili di temperatura mostrano una correlazione positiva con la PC1, mentre quelle di piovosità sono inversamente correlate. Come in Figura 18, BIO₃ e BIO₁₅ sono in netto contrasto rispetto alle loro variabili di appartenenza.

Con la PC2 (11% della varianza) è possibile apprezzare la distribuzione dei campioni secondo le variabili di piovosità BIO₁₂, BIO₁₃, BIO₁₆ e BIO₁₉, che risultano con essa positivamente correlate. L'ultima delle componenti principali (2% della varianza) rappresenta principalmente BIO₃, BIO₁₅ e BIO₄, BIO₇; le prime due variabili risultano essere negativamente correlate con la PC3 mentre le altre due sono positivamente correlate.

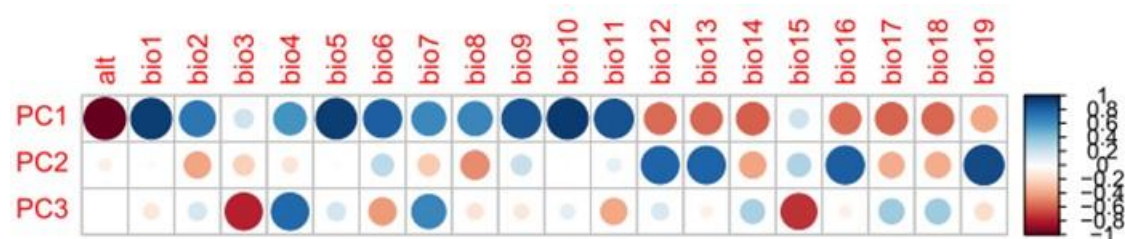


Figura 20. Immagine che rappresenta la correlazione tra le variabili bioclimatiche (BIO₁-BIO₁₉) e l'altitudine (alt) e le tre componenti principali (PC1-PC2-PC3) derivanti dall'analisi di PCA. La dimensione dei cerchi indica il grado di correlazione, grandi dimensioni forte correlazione; il colore dei cerchi indica in che modo sono correlate due variabili, blu positivamente e rosso negativamente in accordo con la scala colore sulla destra.

4.2 Risultati molecolari

L'analisi di RADSeq sul *subset* di 24 campioni (Tabella 4) ha prodotto 864.754 SNP grezzi distribuiti in tutto il genoma di *T. urartu*. Questo numero molto ampio è stato notevolmente ridotto nelle fasi esplorative di cui si occupa questa tesi. Infatti, più che non

il numero totale di marcatori molecolari prodotti, eravamo interessati in questa fase ad identificare un *set* di SNP molto solidi per condurre le prime analisi di diversità. Per questo motivo, i marcatori sviluppati sono stati filtrati per diversi criteri stringenti, tra i quali il numero di *reads* minime necessarie per chiamare uno SNP, il numero massimo di campioni falliti per marcatore, l'assenza di loci eterozigoti e la presenza di un solo polimorfismo per *read* allineata. Il processo di filtraggio ha ridotto il numero di SNP a 8,309. Per la natura dell'approccio molecolare, tutti i loci sono polimorfici per definizione.

Tabella 4. Tabella con il *subset* di 24 campioni, per ogni accessione è riportato il nome originale, la collezione di provenienza, il paese di origine e il *gazetter* dove presente.

Campioni	Seed bag	Collezione	Paese	Gazetteer
8	ID1391	SAL	ARM	NA
12	IDS1541	SAL	ARM	NA
27	ID393	SAL	IRN	6 km S Shahabad, Bakhtaran
40	IDS1535	SAL	IRQ	4 km NE Shaqlawa
42	IDS1538	SAL	IRQ	1 km NE Salahadin
43	IDS1542	SAL	IRQ	Arbil, 1 km NE Salahadin, 1110 m
115	IG 115814	USDA	JOR	Rashidiye old site A
117	IG 115817	USDA	JOR	Ajlun district, Ibbien
120	ID388	SAL	LBN	Talia
122	IDS1529	SAL	LBN	NA
249	IDS1532	SAL	SYR	Aleppo
265	IG110766	USDA	SYR	6 km E Amudah road to Kamishly 3km before Omarabia
299	IDS1534	SAL	TUR	Sanli Urfa
313	IDS1554	SAL	TUR	Mardin, 44.5 km W Kizieltepe, 600 m
324	G1812	USDA	TUR	45.9 km west of Kiziltepe
331	IDS1561	SAL	TUR	Mardin, 53.8 km W Kizieltepe
336	G1827	USDA	TUR	53.8 km west of Kiziltepe
340	G1830	USDA	TUR	53.8 km west of Kiziltepe
341	G1831	USDA	TUR	53.8 km west of Kiziltepe
347	G1839	USDA	TUR	53.8 km west of Kiziltepe
351	IDS1566	SAL	TUR	Mardin, 53.8 km W Kizieltepe
354	G1868	USDA	TUR	83.8 km west of Kiziltepe
332a	G1824	USDA	TUR	53.8 km west of Kiziltepe
332b	G1824	USDA	TUR	53.8 km west of Kiziltepe

A partire da questi dati molecolari è stato costruito un albero filogenetico in *Neighbor-Joining* (Figura 21). La filogenesi ottenuta può essere divisa in almeno quattro gruppi, ove non tutti i campioni sono raggruppati in base ai siti di campionamento. Solo gli individui provenienti dalla Turchia e dalla Giordania sono infatti ordinati in *cluster* monofiletici (gruppi B e C in Figura 21) in accordo alla loro origine geografica. I campioni 8 (Armenia), 40 (Iraq) e 120 (Libano) appartenenti al gruppo A si trovano a essere raggruppati insieme, nonostante la grande distanza geografica tra loro. Questo gruppo di individui appare essere molto distante dal *cluster* principale. Infine, l'albero evidenzia la presenza di un gruppo nel quale sono raggruppate diverse accessioni (gruppo D). In questo gruppo, la diversità genetica non è ordinata secondo la provenienza geografica delle accessioni. I campioni presenti, infatti, provengono da differenti paesi anche molto distanti tra loro: Armenia (campione 12), Iran (27), Iraq (42, 43), Siria (249, 265), e Turchia (299).

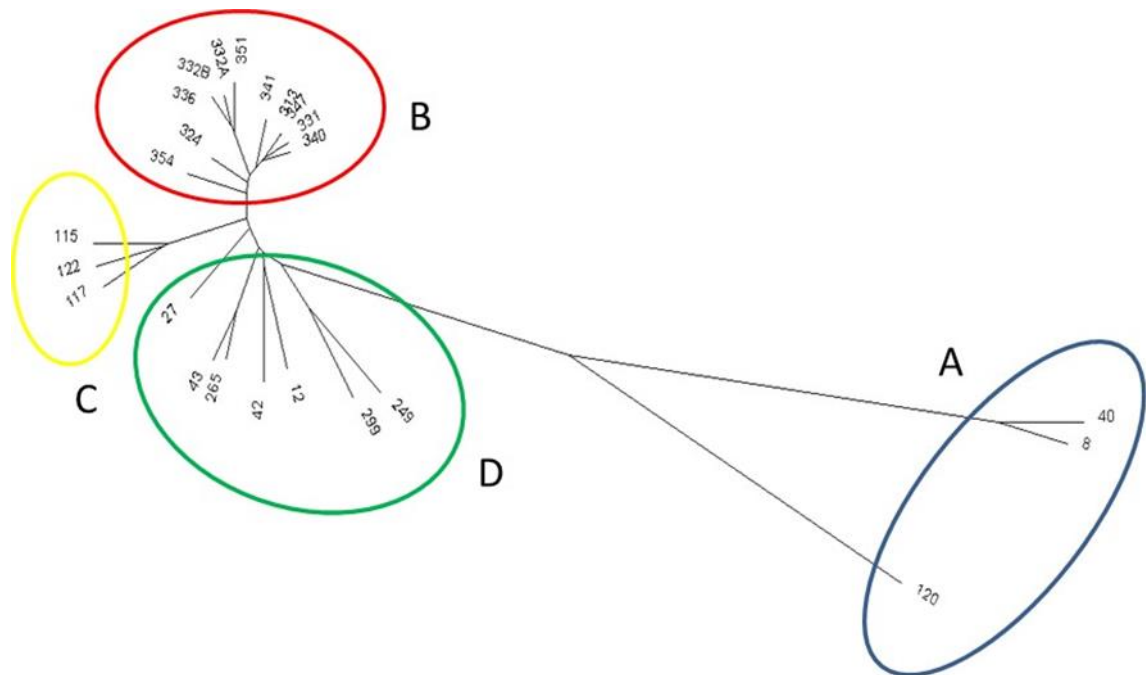


Figura 21. Analisi filogenetica effettuata sul *subset* di 24 campioni che indica la diversità molecolare all'interno della collezione. L'albero risultante è diviso in 4 gruppi (A, B, C, D).

Così come la PCA delle variabili bioclimatiche suggeriva grande commistione tra le diverse nazioni di campionamento (Figura 19) anche l'analisi filogenetica, sebbene effettuata su un piccolo *subset*, suggerisce un'elevata variabilità tra le accessioni della collezione. Di nuovo, la variabilità molecolare non è distribuita secondo i paesi di origine (Figura 22). Osservando l'asse di PC1 della PCA sul dato molecolare è possibile notare come, anche in questo caso, i campioni 8, 40 e 120 si trovino nettamente separati dal resto delle accessioni (in accordo con l'albero filogenetico di Figura 21). Queste tre accessioni sono infatti distribuite a valori estremamente positivi di PC1 rispetto al resto dei campioni che si trovano principalmente a valori negativi.

I campioni turchi (punti gialli su valori di PC2 positivi) risultano essere molto simili tra loro sull'asse di PC1, mentre mostrano differenze sull'asse di PC2. Questi risultati sono in accordo con l'analisi filogenetica (Figura 21): le accessioni provenienti

da una stessa regione, ad esempio i punti gialli per i campioni turchi, risultano riunite in un *cluster* molto stretto sulla PC1, ma sono maggiormente differenziati dalla PC2. Questa separazione interna è confermata dalla struttura dei nodi alla base dei cluster prodotti dall'albero filogenetico. Come nell'albero filogenetico, troviamo, il gruppo contenente le accessioni giordane vicino ai campioni turchi. A valori negativi di PC2 si trovano tutti i campioni appartenenti al gruppo D (Figura 22), quello composto da accessioni con siti di campionamento geograficamente distanti (Armenia, Iran, Iraq, Siria e Turchia).

E' importante notare come la varianza spiegata dai due primi assi di PCA raggiunga solo il 48% (rispettivamente il 38% per PC1 ed il 10% per PC2). La bassa varianza spiegata dagli assi principali suggerisce che la diversità molecolare nel *dataset* è scarsamente strutturata, e difficilmente può essere condensata in un numero limitato di variabili PC.

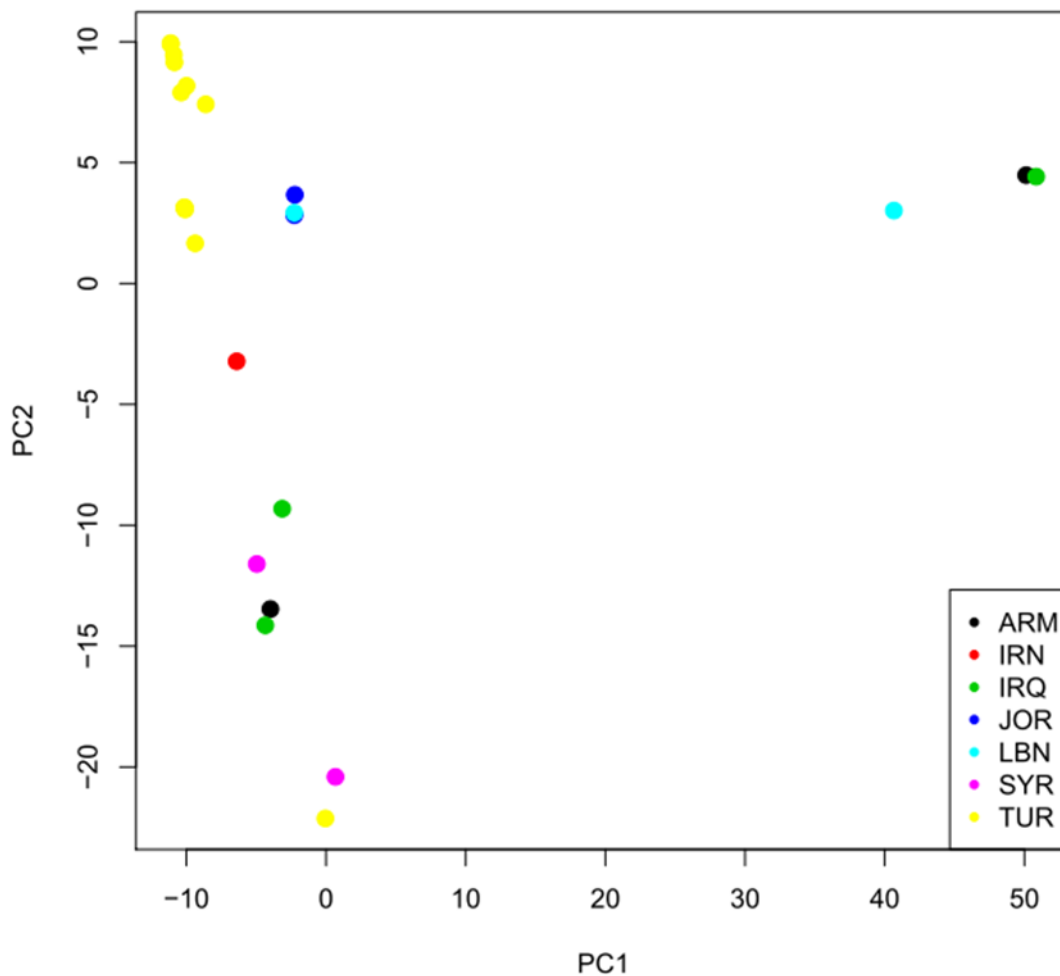


Figura 22. PCA dei dati molecolari ottenuti dal RAD Seq del *subset* di 24 campioni. L'immagine rappresenta le due componenti principali che spiegano rispettivamente il 38% e il 10% della varianza del *dataset*, la legenda in basso a destra indica la provenienza dei campioni, a ogni regione appartiene un colore diverso (ARM= Armenia, IRN = Iran, IRQ = Iraq, JOR = Giordania, LBN = Libano, SYR = Siria, TUR = Turchia).

4.3 Risultati fenotipici

4.3.1 Il campo sperimentale

Intorno a metà Aprile il campo si presentava come in Figura 23, con tutte le accessione della collezione germogliate e in piena crescita. La velocità di accestimento e crescita si

è presentata da subito molto variabile per le diverse accessioni, indicando un potenziale vegetativo molto diverso.



Figura 23. Foto del campo di *Triticum urartu* di Aprile 2015. Notare i diversi gradi di accostimento e crescita per le diverse accessioni. Centro di Ricerca Centro Interuniversitario sulle Biomasse da Energia (CRIBE) in San Piero a Grado (Pisa, Italia).

Al termine di Aprile solo le accessioni più precoci avevano iniziato a spigare (Figura 24a), mentre la maggior parte della collezione risultava essere ancora nello stadio fenologico di botticella. In questa fase sono apparsi elementi tipici che caratterizzano le specie selvatiche, tra questi il nodo pubescente e il colore violaceo della foglia bandiera (Figura 24b-c).



Figura 24. Foto di alcune accessioni di *Triticum urartu* il 24 Aprile 2015: **a)** Particolare della prima spiga; **b)** Particolare del nodo pubescente tipico delle specie selvatiche; **c)** Particolare della lamina fogliare della foglia bandiera, il colore violaceo è tipico delle specie selvatiche. Centro di Ricerca Centro Interuniversitario sulle Biomasse da Energia (CRIBE) in San Piero a Grado (Pisa, Italia).

Nelle prime settimane di Maggio quasi tutte le accessioni erano spigate (Figura 25 a-b).

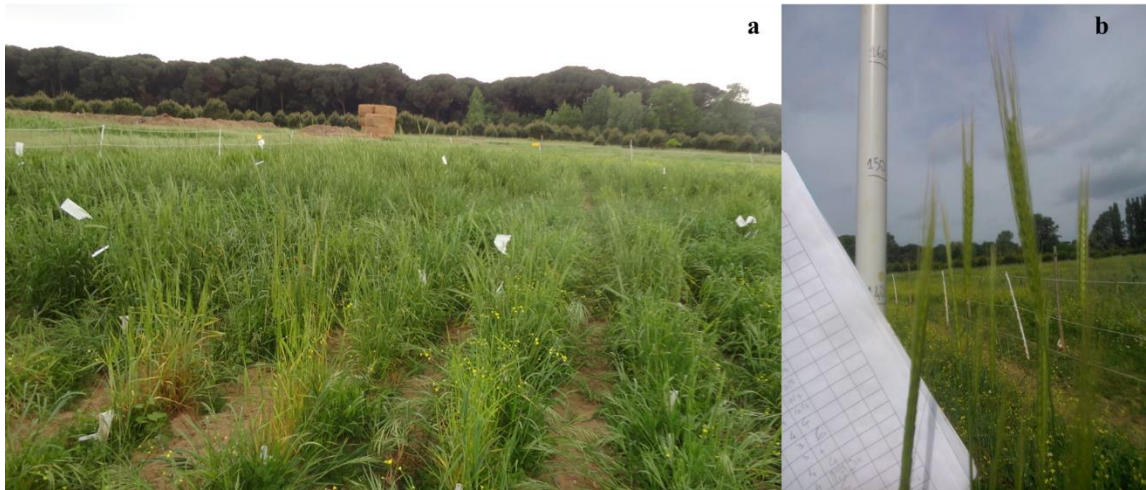


Figura 25. Foto del campo di *Triticum urartu* delle prime settimane di Maggio 2015: **a**) Campo di maggio dove quasi tutte le accessioni sono spigate, i sacchetti bianchi indicano gli incroci effettuati tra *T. urartu* sequenziato (G1812) e altre accessioni; **b**) Particolare dei rilievi effettuati per i caratteri fenotipici, in questo caso l'altezza della pianta. Centro di Ricerca Centro Interuniversitario sulle Biomasse da Energia (CRIBE) in San Piero a Grado (Pisa, Italia).

Nelle ultime settimane di Maggio anche le accessioni più tardive hanno spigato, e le spighe hanno iniziato a maturare come indicato dal loro cambio di colore da verde a giallo-bruno (Figura 26).



Figura 26. Foto del campo di *Triticum urartu* delle ultime settimane di Maggio 2015: le accessioni sono tutte spigate ed hanno quasi raggiunto la maturità. Le corde sono necessarie per evitare l'allettamento delle piante. Centro di Ricerca Centro Interuniversitario sulle Biomasse da Energia (CRIBE) in San Piero a Grado (Pisa, Italia).

A maturità ultimata, è stato di nuovo possibile osservare una buona variabilità all'interno della collezione. Accessioni provenienti dalla medesima regione mostravano spighe decisamente diverse, e caratteristiche morfologiche divergenti (Figura 27).



Figura 27. Foto del campo di *Triticum urartu* delle ultime settimane di Maggio 2015: Particolare di due spighe libanesi provenienti da accessioni differenti, si apprezza la diversità presente all'interno della stessa regione. Centro di Ricerca Centro Interuniversitario sulle Biomasse da Energia (CRIBE) in San Piero a Grado (Pisa, Italia).

4.3.2 Risultati dati fenotipici

La distribuzione di sei degli otto caratteri fenotipici è stata osservata utilizzando il programma R. Poiché i dati sulla presenza o assenza delle reste e sulla pelosità della spiga erano a risposta binaria, cioè affermativa o negativa, questi dati sono stati esclusi dalle analisi.

Dalle analisi condotte sono stati prodotti sei istogrammi rappresentativi dei caratteri fenotipici. La distribuzione dell'altezza delle piante (Figura 28a) presenta un andamento normale, cioè la maggioranza degli individui ha altezze medie comprese tra 120-140 cm (media a $127,5 \pm 15,3$ cm). Il resto delle accessioni è concentrato soprattutto in fenotipi bassi (60-120 cm), con pochi individui con altezze superiori ai 140 cm. L'epoca di spigatura (Figura 28b), invece, si discosta maggiormente da una distribuzione perfettamente normale, e mostra un profilo bipartito. La gran parte delle accessioni è

spigata intorno ai primi di maggio, ma il *range* dell'epoca di fioritura risulta straordinariamente ampio e supera le tre settimane.

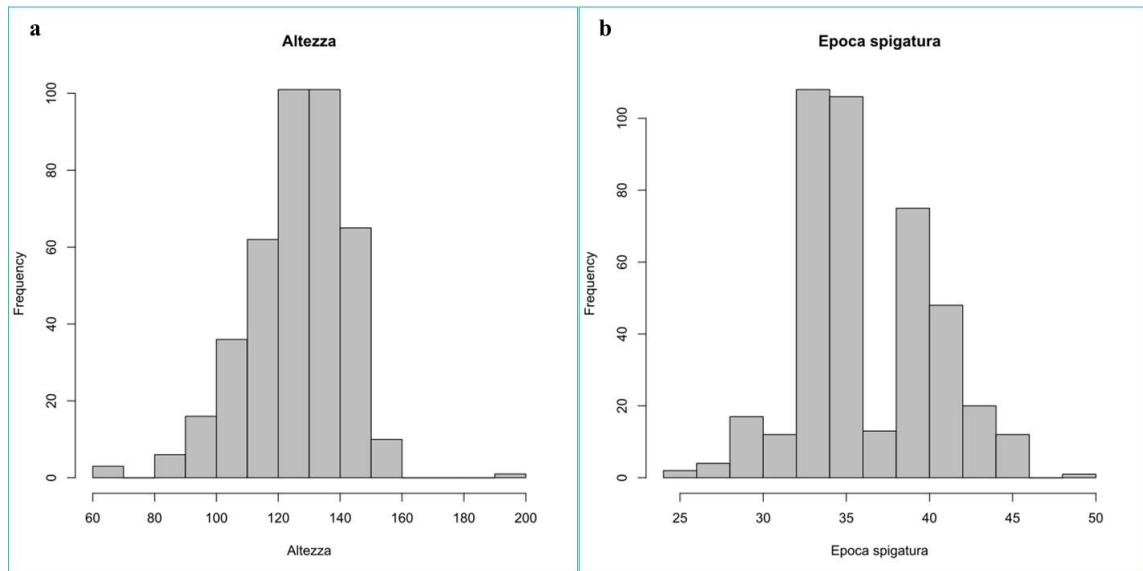


Figura 28. Immagine che rappresenta due istogrammi: **a)** Altezza, sull'ascissa si trovano le altezze delle piante espresse in cm; **b)** Epoca di spigatura, sull'ascissa si trova la data di spigatura delle accessioni espressa in giorni. In entrambi i grafici sulle ordinate troviamo la distribuzione dei campioni.

Nel grafico del portamento (Figura 29a) è possibile notare che i campioni si trovano maggiormente distribuiti tra i valori 2-3-4, indicando che un elevato numero di accessioni presenta un portamento eretto. Esistono poi diversi campioni con portamento prostrato (valore 1) e in misura minore con un portamento perfettamente eretto. La maggior parte dei campioni presenta un'area fogliare compresa tra 5 e 20 cm² (media a $11,8 \pm 6,6$ cm²), e solo poche accessioni presentano foglie bandiera significativamente più grandi (Figura 29b).

In Figura 30 sono riportati gli istogrammi della lunghezza della spiga e del numero di spigette. Anche per la lunghezza delle spighe (Figura 30a) la distribuzione dei campioni ha un andamento normale ($10,6 \pm 1,7$ cm). Le code di questo grafico in cui

troviamo i valori estremi di lunghezza presentano invece poche accessioni. La collezione presenta un numero di spighe medio di $25,9 \pm 5,6$ (Figura 30b).

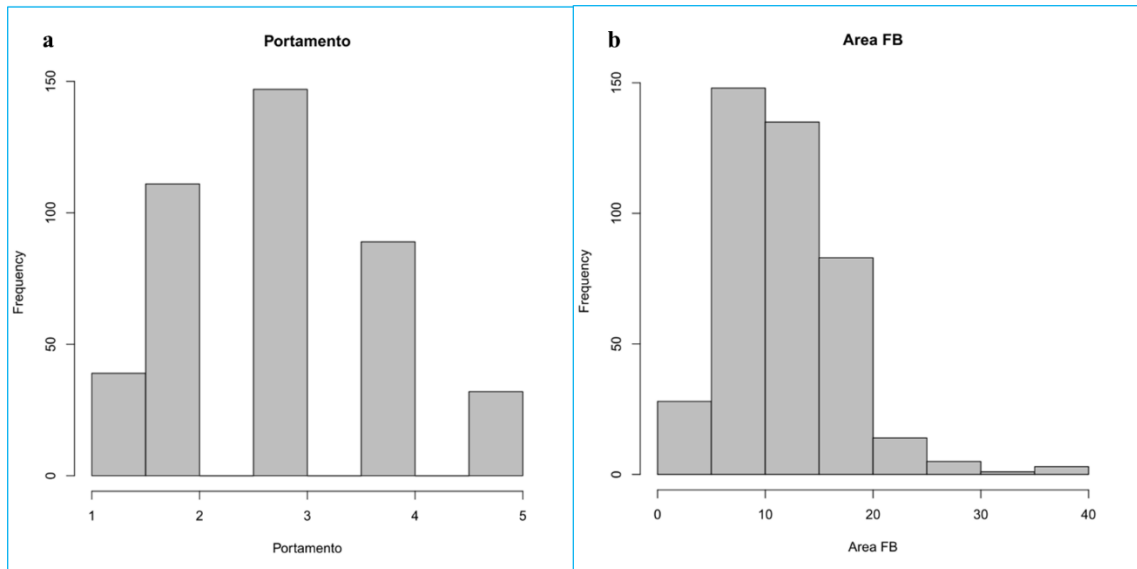


Figura 29. Immagine che rappresenta due istogrammi: **a)** Portamento, sull'ascissa si trova il portamento delle piante espresso con valori da 1, prostrato, a 5, eretto; **b)** Area fogliare, sull'ascissa l'area della foglia bandiera espressa in cm^2 . In entrambi i grafici sulle ordinate troviamo la distribuzione dei campioni.

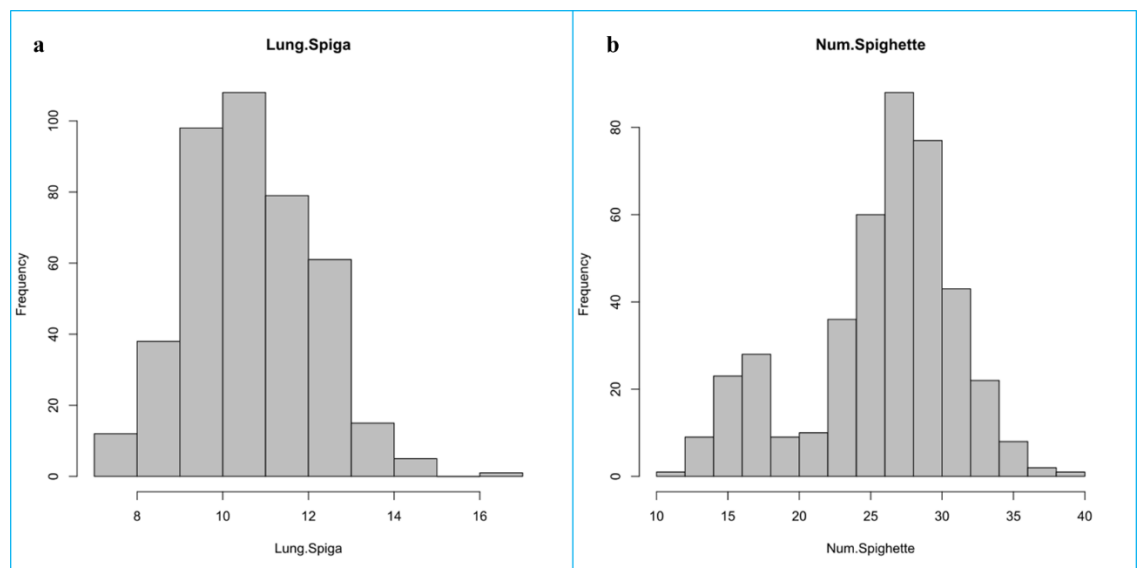


Figura 30. Immagine che rappresenta due istogrammi: **a)** Lunghezza della spiga, sull'ascissa è riportata la lunghezza espressa in cm; **b)** Numero di spighette per spiga, sull'ascissa è riportato il numero delle spighette. In entrambi i grafici sulle ordinate troviamo la distribuzione dei campioni.

Il grado di correlazione tra i caratteri fenotipici (Figura 31) ha evidenziato che il portamento e l'altezza della pianta risultano essere negativamente correlate con l'epoca di spigatura. L'altezza della pianta mostra anche una leggera correlazione con la lunghezza della spiga e il numero di spighe. La lunghezza della spiga e il numero delle spighe come prevedibile risultano tra loro positivamente correlate, mentre non mostrano correlazione con l'area della foglia bandiera.

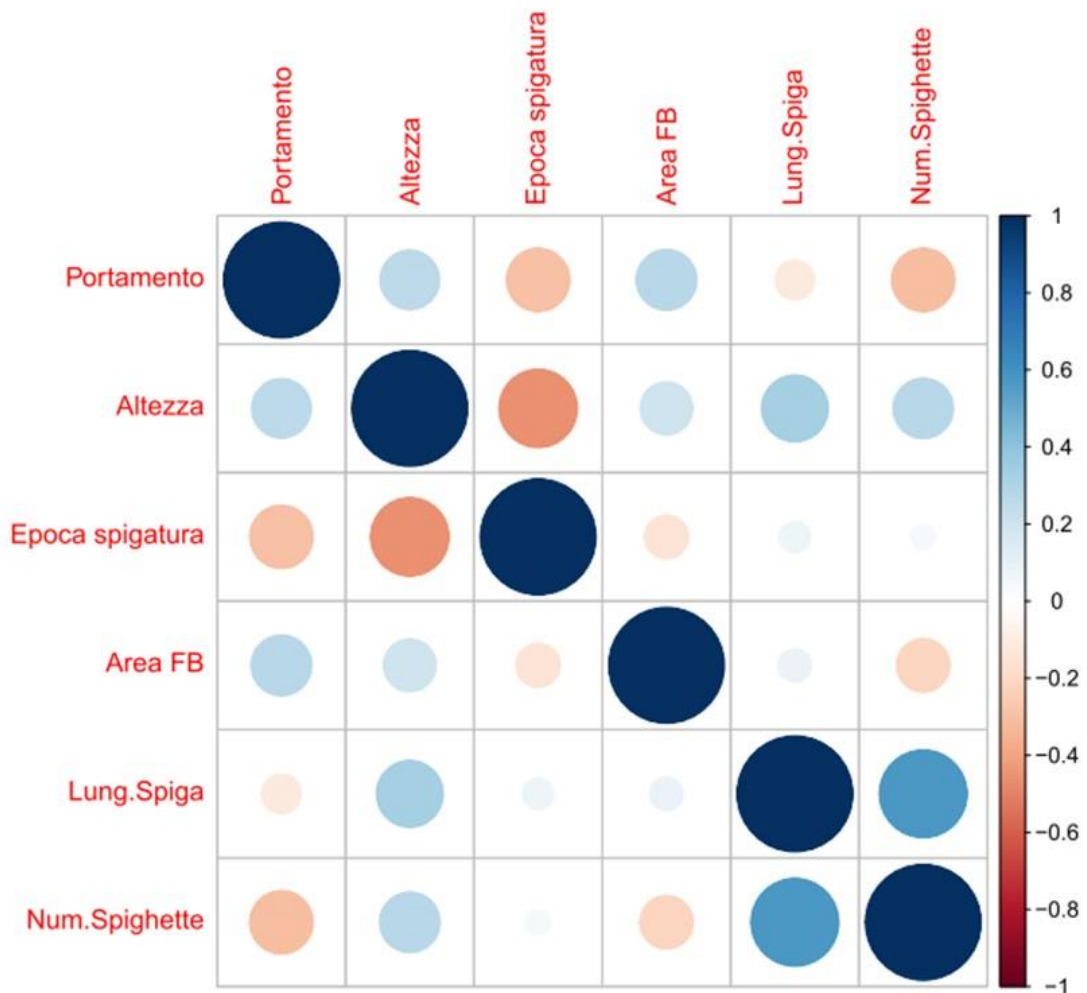


Figura 31. Immagine che rappresenta la correlazione tra i caratteri fenotipici raccolti, sulla destra si trova una scala che indica il grado di correlazione. La dimensione dei cerchi indica il grado di correlazione, grandi dimensioni forte correlazione; il colore dei cerchi indica in che modo sono correlate due variabili, blu positivamente e rosso negativamente. I caratteri fenotipici sono confrontati contro loro stessi.

Anche l'analisi di PCA dei caratteri fenotipici secondo le nazioni di campionamento propone dei risultati interessanti (Figura 32). PC1 e PC2 spiegano rispettivamente il 78% e l'11% della varianza del *dataset*. Rispetto alla PCA sui dati ecologici non si apprezza una netta divisione, ma è possibile osservare che individui

appartenenti alla stessa regione tendono a raggrupparsi tra loro. La maggior parte dei campioni iracheni (punti verdi) si trovano principalmente su valori di PC2 positivi, mentre la maggioranza di quelli libanesi (punti azzurri) sono distribuiti a valori di PC2 negativi. Quasi tutti i campioni armeni risultano essere distribuiti tra valori di PC2 molto simili tra loro, a differenza di quelli provenienti dall'Iran che risultano essere ampiamente distribuiti su diversi valori di PC2. Le accessioni turche (punti grigi) e in misura minore quelle siriane (punti gialli) risultano essere simili tra di loro sulla PC1 mentre sull'asse di PC2 si apprezza un'elevata distribuzione.

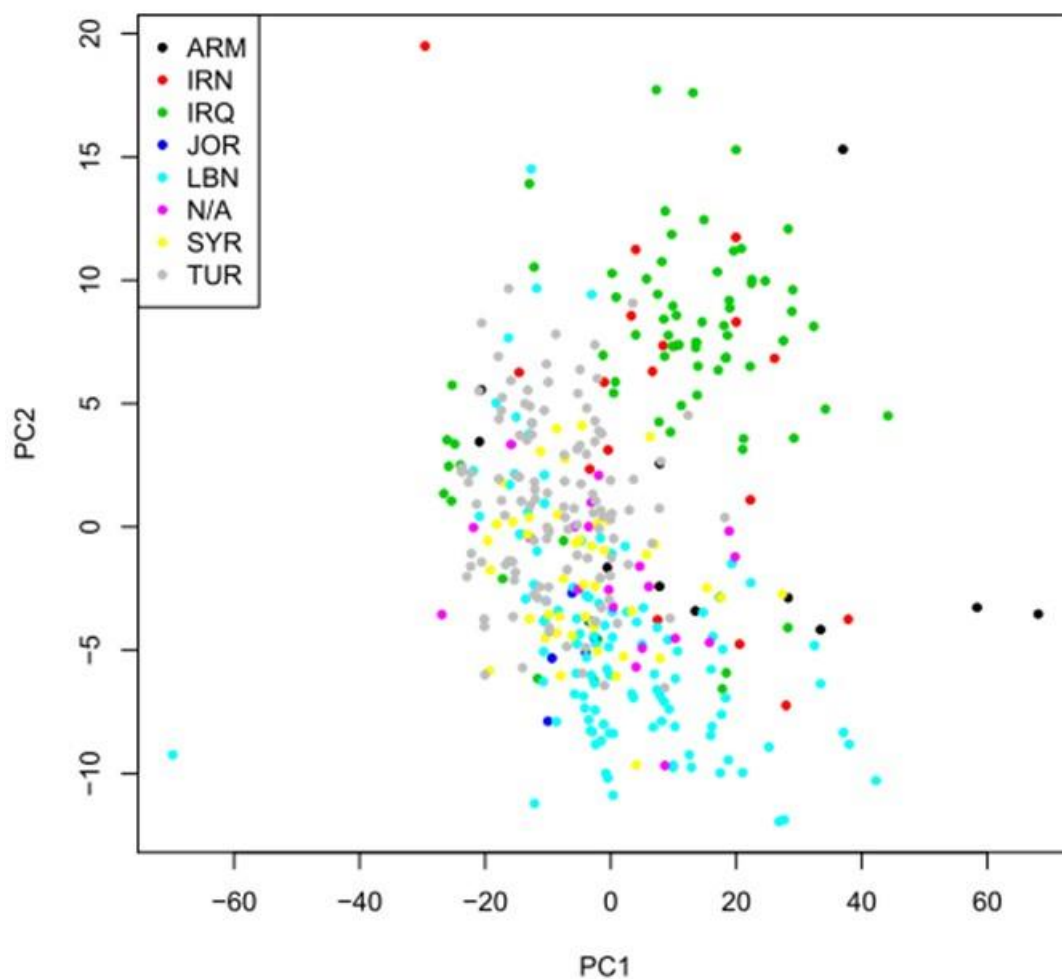


Figura 32. PCA dei caratteri fenotipici secondo i siti di campionamento delle accessioni. L'immagine rappresenta le due componenti principali che spiegano rispettivamente il 78% e il 11% dei campioni, la legenda in alto a sinistra indica la provenienza dei campioni, a ogni regione appartiene un colore diverso (ARM= Armenia, IRN = Iran, IRQ = Iraq, JOR = Giordania, LBN = Libano, SYR = Siria, TUR = Turchia).

Con lo scopo di comprendere quali fossero le variabili originali ad avere il maggior peso nella determinazione dei primi tre assi di PC, è stata compiuta un'analisi di correlazione (Figura 33). La PC1, che spiega il 78% della varianza del *dataset*, è soprattutto contribuita dall'altezza e in misura minore dagli altri caratteri fenotipici.

L'altezza della pianta è inversamente correlata con la prima PC, ciò significa che all'aumentare dei valori di PC1 l'altezza diminuisce.

Con la PC2 (11% della varianza) è possibile apprezzare la distribuzione dei campioni secondo il numero di spighe, con cui mostra una correlazione negativa, e l'area fogliare con la quale risulta essere positivamente correlata; e in misura minore il portamento (positivamente correlata), l'epoca di spigatura e la lunghezza della spiga (negativamente correlate con entrambe). La PC3 (6% della varianza) risulta essere positivamente correlata con l'area della foglia bandiera e il numero di spighe, e in misura minore con la lunghezza della spiga.

I campioni armeni presentano altezze molto differenti tra loro, infatti risultano essere ampiamente distribuiti lungo l'asse di PC1, mentre presentano l'area della foglia bandiera e il numero di spighe molto simili (Figura 32). Considerando la distribuzione secondo l'asse di PC2 si apprezza una notevole variabilità in tutte le accessioni, indice di notevoli differenze nei valori dell'area della foglia bandiera e del numero spighe. I due gruppi che presentano una distribuzione maggiore lungo questo asse sono costituiti da accessioni turche e libanesi. La variabilità in questi due gruppi è evidenziata dal range di valori di PC2 in cui si trovano

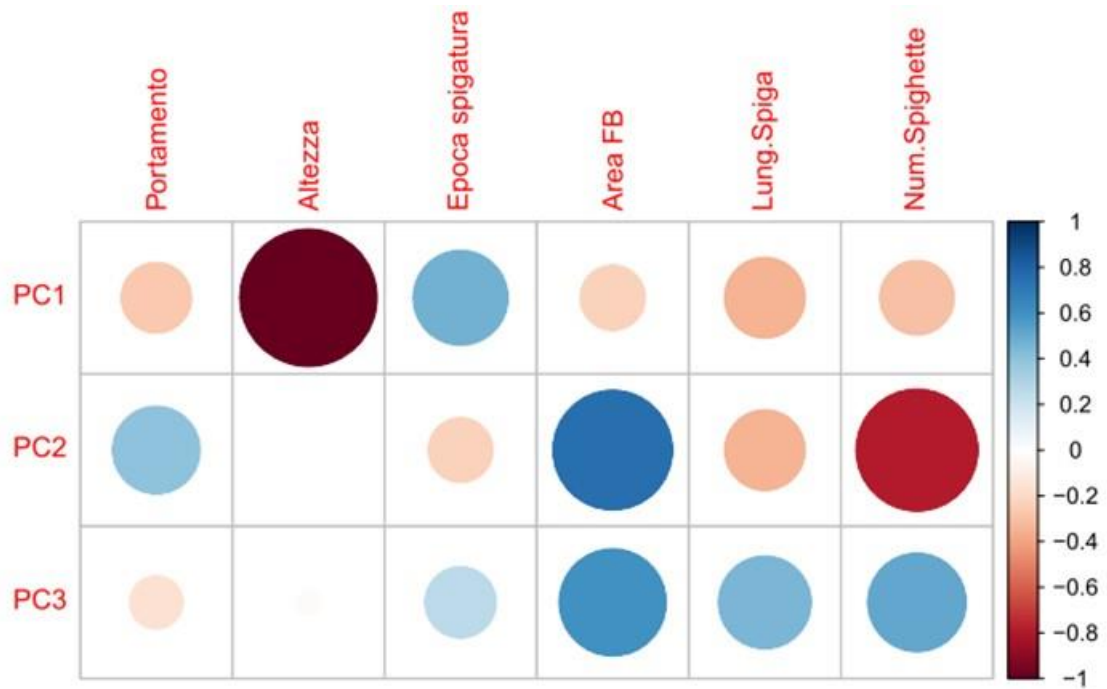


Figura 33. Immagine che rappresenta la correlazione tra i caratteri fenotipici raccolti e le tre componenti principali (PC1-PC2-PC3), sulla destra si trova una scala che indica il grado di correlazione. La dimensione dei cerchi indica il grado di correlazione, grandi dimensioni forte correlazione; il colore dei pallini indica in che modo sono correlate due variabili, blu positivamente e rosso negativamente. Sulla destra si trova una scala che indica il grado di correlazione tra le variabili

5 Discussione

5.1 Distribuzione ecologica di *Triticum urartu*

Il territorio della Mezzaluna Fertile è estremamente vario da un punto di vista orografico: vi si possono trovare montagne alte fino a 5000m, pianure e la depressione più bassa al mondo, il Mar Morto (Diamond 2015). Da tale ricchezza ambientale era possibile aspettarsi un'elevata variabilità in termini di distribuzione delle popolazioni di *T. urartu*, variabilità che è stata riscontrata nelle analisi ecologiche e geografiche condotte sulla collezione oggetto di questa tesi.

Proiettando i siti di campionamento su una mappa altitudinale (Figura 12), è stato possibile vedere come le accessioni incluse nella collezione presentassero uno spettro ampio di valori altitudinali. Le accessioni infatti provengono da un *range* altitudinale di più di 2000 m, variando da un minimo di 50 ad un massimo di 2367 m. Bisogna sottolineare che l'area di campionamento attraversa diverse latitudini e zone climatiche: per questo motivo, è possibile trovare popolazione di *T. urartu* a così diversi valori altitudinali. Prevedibilmente, la temperatura (Figura 13) e la piovosità (Figura 14) nell'area di campionamento seguono un simile *range* di variazione, dalle zone basse, calde e aride a quelle elevate, più fresche e piovose. Come discusso in seguito, la covariazione di queste variabili non è sempre lineare proprio a causa dell'estensione dell'area di campionamento.

Nonostante l'estensione del campionamento, la distribuzione dei campioni secondo l'altitudine non presenta un andamento normale (Figura 17). Questo è dovuto al fatto che i campionamenti originali sono stati condotti in diverse occasioni e in studi

indipendenti, non tracciando un transetto lungo tutte le migliaia di chilometri che abbraccia la collezione. Questo ha fatto sì che i campioni fossero concentrati attorno alle aree di campionamento, e non seguono l'intero spettro di diffusione di *T. urartu*.

Le mappe geografiche della temperatura e della piovosità media annuale (Figura 13,14) così come quella dell'altitudine (Figura 12) forniscono un'informazione relativa al *trend* generale di crescita di *T. urartu*. Accoppiando questi dati con le altre variabili di *Bioclim* è possibile ottenere una visione più ampia e soprattutto un livello di informazione più dettagliato. Osservando dunque le altre variabili legate alla temperatura (Figura 15) viene subito messo in risalto l'importanza di questi dati. La crescita di una pianta infatti non è legata alla sola temperatura media annuale ma anche, per esempio, alla temperatura massima e minima durante il suo ciclo vitale. Con il solo dato della temperatura media annuale, si potrebbe formulare l'ipotesi che l'*optimum* di crescita delle accessioni sia compreso tra un massimo di 23° C e un minimo di 4°C, quando in realtà è riportata l'esistenza di numerosi campioni che sono in grado di sopportare inverni estremamente rigidi con temperature che arrivano anche a -10° C. La piovosità media annuale (Figura 16) presenta una distribuzione più ampia rispetto a quella dei valori di temperatura media annuale: la collezione comprende infatti accessioni che crescono a valori estremi di piovosità compresi tra un minimo di 97 mm e un massimo di 1029 mm. Questo dato è un'ulteriore dimostrazione dell'elevata variabilità ambientale in cui cresce la collezione di *T. urartu*.

Come previsto, dalle analisi è emersa una forte correlazione tra l'altitudine e le principali variabili bioclimatiche. E' invece interessante notare come questa correlazione non risulti essere sempre lineare (Figura 18). BIO₃ e BIO₁₅ sono infatti discordanti rispetto

alle altre variabili. Queste variabili, che riportano misure di stagionalità legate alla temperatura e alla piovosità, indicano con la loro discordanza in rapporto all'altitudine ed alle piovosità e temperature medie la presenza di diverse fasce climatiche nell'area di campionamento di *T. urartu*. Questo risultato fornisce un'ulteriore evidenza dell'importanza di utilizzo di tutte le variabili legate al clima. Questi dati bioclimatici interpolati ottenuti tramite *Bioclim*, sono fondamentali in analisi di *landscape genomics* (Dell'Acqua *et al.* 2014) perché forniscono una visione ampia dello spettro di variabili ecologiche che contribuiscono alle condizioni particolari in cui cresce una determinata specie. Questi dati coprono un arco temporale che va dal 1950 al 2000, e sono quindi centrate sull'epoca dei campionamenti delle accessioni incluse nella collezione, che risalgono agli anni '70. Sarebbe interessante ottenere dati climatici che arrivano fino ai giorni nostri, e questo perché il cambiamento climatico globale si è fatto sempre più pressante negli ultimi 15 anni. In futuro, nuovi dati climatici e pedologici saranno associati ai campioni che compongono la collezione, permettendo un'analisi più fine della variabilità pedoclimatica nei siti di campionamento.

L'effetto delle differenti latitudini attraversate dalle aree di raccolta risulta maggiormente evidente dall'analisi di PCA sui dati bioclimatici (Figura 19). In questa analisi, i campioni raggruppati in base ai dati bioclimatici non mostrano un raggruppamento sulla base delle regioni di campionamento. Osservando la distribuzione dei campioni secondo l'asse di PC1 (86% della varianza) è possibile notare quanto i campioni armeni si discostino dal resto della collezione. I siti di campionamento delle accessioni armenie sono situati infatti in zone che presentano altitudini elevate, questa loro diversità si traduce graficamente con la loro distribuzione a valori di PC1 estremamente negativi. Il resto delle accessioni presenta invece una distribuzione molto ampia lungo

questo asse di PC, indicando cioè che il resto dei campioni cresce attraverso un *range* altitudinale più vasto e quindi che si trova distribuito in un territorio che presenta maggiore variabilità ambientale. È interessante notare inoltre come campioni provenienti dalla stessa regione e che hanno valori uguali sull'asse di PC1, cioè siti di campionamento alla medesima altitudine, presentano valori molto diversi sull'asse di PC2. Questo asse di PC è contribuito maggiormente dalle variabili di piovosità (Figura 20) ciò significa che la localizzazione delle accessioni a valori positivi di PC corrisponde a regioni con regimi intensi di piovosità e viceversa. Le accessioni che crescono, come in questo caso, in regimi di piovosità contrastanti possono essere considerati dei buoni candidati per l'adattamento all'uno o all'altro estremo di questa condizioni.

I risultati dell'analisi geografica dimostrano l'esistenza di un'elevata variabilità ambientale non solo all'interno della collezione ma anche tra accessioni provenienti dalla stessa regione. La distribuzione delle accessioni secondo svariati ambienti ecologici e quindi di diverse fasce climatiche, è indice della capacità di *T. urartu* di adattarsi a differenti condizioni climatiche e quindi alla possibile presenza di alleli adattativi nel suo genoma (Jackson, Ford-Lloyd, e Parry 2013b). Questa diversità è fondamentale per la ricerca di quei caratteri di rusticità che consentono alle popolazioni di *T. urartu* di adattarsi alle differenti condizioni ambientali a cui è sottoposto. La scoperta degli alleli correlati a questa variabilità ambientale, una volta che l'intera collezione sarà caratterizzata, può portare un contributo significativo ai programmi di miglioramento genetico del frumento (Rouse e Jin 2011). Nuovi campionamenti eseguiti allo scopo di riempire i *gap* della collezione corrente saranno estremamente utili all'ulteriore definizione della reale distribuzione di *T. urartu*. Purtroppo, al momento della stesura di questa tesi, le

condizioni politiche dell'areale di crescita di *T. urartu* sono proibitive e impediscono l'organizzazione di spedizioni di raccolta nel breve e medio termine.

5.2 Diversità molecolare in *Triticum urartu*

Le 24 accessioni incluse nel *subset* genotipizzato hanno prodotto più di 860,000 marcatori molecolari. Sebbene non tutti questi SNP passerebbero i controlli qualità necessari alle analisi di diversità molecolare, il loro numero rappresenta un ottimo punto di partenza per garantire una rappresentazione dell'intero genoma di *T. urartu*. Considerando una dimensione del genoma di *T. urartu* di 5 Gb (Ling *et al.* 2013), questo porterebbe ad uno SNP ogni 5.800 nucleotidi. Questo numero è certamente sovrastimato dalla presenza di marcatori erroneamente identificati come tali, a causa di errori di sequenziamento ed errori di allineamento in sede di chiamata della variabilità allelica. Analisi preliminari sul dato molecolare, non incluse in questa tesi, riportano in 20.000-30.000 il numero di SNP utilizzabili una volta caratterizzata l'intera collezione. Questo numero porterebbe comunque ad uno SNP ogni 200.000 nucleotidi, una densità di genotipizzazione mai raggiunta in *T. urartu* ed in studi sui progenitori selvatici del grano.

La tecnica di RADseq utilizzata si è quindi dimostrata adatta alla caratterizzazione della diversità presente nella collezione. Le accessioni genotipizzate in questa fase sono state selezionate con lo scopo di osservare il limite superiore e quello inferiore della diversità genetica inclusa nella popolazione. In questa fase sono stati utilizzati dei criteri prudenziali per ridurre al minimo l'eventualità di un allineamento fallace delle *reads* e di conseguenza l'identificazione di SNP artificiali. Sono stati esclusi tutti gli SNP eterozigoti poiché l'eterozigosità attesa di *T. urartu*, una specie autogama, è molto bassa.

In questo contesto gli SNP eterozigoti potevano derivare da un allineamento delle *reads* su sequenze ripetute, e quindi su più loci nel genoma, originando eterozigotità apparente.

Selezionando quindi solo gli SNP più solidi (1/100 di quelli prodotti), è stato possibile dare una prima valutazione della variabilità genetica presente nel *subset*. I risultati dell'analisi filogenetica e della PCA (Figura 21, 22) evidenziano come la provenienza geografica non sia sempre in accordo con la diversità genetica, una condizione già osservata in popolazioni naturali di *T. urartu* tramite l'uso di isoenzimi (Moghaddam, Ehdaie, e Waines 2000). La distanza genetica dei campioni 8 (Armenia) e 40 (Iraq) appartenenti al gruppo A, rispetto a tutte le altre accessioni è probabilmente causata da differenze tassonomiche tra questi campioni e le accessioni di *T. urartu sensu stricto*. In particolare, l'accessione 8 è nota appartenere alla subsp. *album*, mentre il campione 40 potrebbe addirittura essere *T. monococcum* subsp. *beoticum*. Il campione 120, ugualmente separato dal gruppo principale, non è al momento caratterizzato come sottospecie di *T. urartu*. La presenza di sottospecie può essere giustificata dal continuo flusso genico che spontaneamente avviene in natura tra *T. urartu* e *T. monococcum*. Studi recenti hanno quantificato l'interfertilità tra queste due specie, riscontrando la presenza di ibridi naturali tra le due (Fricano *et al.* 2014). Questa diversità potrebbe essere alla base della forte separazione del gruppo A. L'estensione del numero di accessioni genotipizzate permetterà di quantificare questo fenomeno in maniera più puntuale.

E' interessante notare che, mentre i campioni turchi si raggruppano in un *cluster* monofiletico (Figura 21), i campioni appartenenti al gruppo D mantengono una certa similarità a livello molecolare pur provenendo da siti di campionamento molto distanti tra loro (12 Armenia; 27 Iran; 42, 43 Iraq; 249, 265 Siria e 299 Turchia). Questo significa

che la distanza geografica non è il solo, o il principale, fattore che contribuisce al differenziamento molecolare. Il modello di isolamento da distanza (Wright 1943) non sembra da questi dati preliminari descrivere la condizione delle popolazioni naturali di *T. urartu* attraverso la Mezzaluna Fertile. L'estensione delle analisi molecolari all'intera collezione permetterà una caratterizzazione più fine della distribuzione della diversità genetica in relazione alla diversità geografica.

Altre popolazioni naturali di monocotiledoni hanno mostrato una simile discrepanza con il modello di isolamento da distanza (Dell'Acqua *et al.* 2014). Ci sono diverse possibili spiegazioni alla formazione di gruppi monofiletici di accessioni così distanti da un punto di vista geografico. La forza alla base di questo raggruppamento potrebbe essere la migrazione: è possibile che alcune accessioni abbiano allargato il loro areale di distribuzione andando ad amalgamarsi con popolazioni già esistenti e costituendo infine un gruppo eterogeneo formato da campioni provenienti da zone anche geograficamente molto distanti. La vicinanza genetica che caratterizza il gruppo D può anche dipendere dalla capacità di adattamento di queste accessioni a determinate condizioni ambientali. Nonostante, infatti, queste accessioni siano state campionate in regioni diverse è possibile immaginare che in queste aree siano presenti condizioni climatiche simili che hanno dunque portato alla selezione di una determinata linea genetica. Questo è il tratto di maggiore interesse per la *landscape genomics*, in quanto è proprio l'effetto della selezione naturale l'obiettivo ultimo di questi approcci. Identificare loci che siano soggetti a selezione naturale, o linee genetiche adattate a peculiari condizioni ambientali, è il prerequisito allo sfruttamento di popolazioni naturali di *T. urartu* nel miglioramento del frumento coltivato.

All'epoca della tesi il numero esiguo di campioni permette solo di formulare ipotesi in questo senso: la scelta dei campioni da analizzare non era tanto mirata alla rappresentazione di tutto l'areale di raccolta quanto alla valutazione della diversità attesa in accessioni vicine e distanti. Per questo motivo, il fatto che esista una grande distanza geografica tra linee genetiche simili può essere semplicemente causata dal cosiddetto errore di campionamento, una forma di deriva genetica o *drift*, derivato dal numero esiguo di campioni selezionati: è possibile che linee genetiche diverse presenti nello stesso sito di campionamento non siano state selezionate nell'analisi, e quindi non appaiano nel risultato finale fornendo una visione parziale della distribuzione della diversità di *T. urartu* attraverso la Mezzaluna Fertile. Grazie al sequenziamento dell'intera collezione sarà possibile dare una valutazione statistica alle varie variabili in gioco e cercare così di ridurre al minimo l'ipotesi dell'errore di campionamento. Valutando in maniera quantitativa migrazione, *drift*, e selezione naturale sarà possibile identificare le accessioni ed i loci di maggiore interesse per lo sfruttamento dell'adattamento di popolazioni naturali di *T. urartu*.

5.3 Caratterizzazione fenotipica

L'elevata variabilità presente all'interno della collezione è stata ulteriormente dimostrata dai rilievi fenotipici effettuati in campo. Sebbene la differenza riscontrata sulla velocità di crescita e accostamento tra le accessioni è apparsa subito evidente (Figura 23), questa può essere contribuita dalla differente densità di semina. Allo scopo di produrre quanto più seme possibile, infatti, tutto il materiale disponibile per ciascuna accessione è stato seminato. Un'ulteriore variabile in questo caso è quella della diversa provenienza e conservazione dei semi, che può avere effetti sulla germinabilità in campo. E' noto che

un numero maggiore di semi crea una maggiore competizione tra le piante in crescita andando quindi ad aumentarne l'accestimento (Valério *et al.* 2013). Questo tipo di carattere può essere particolarmente influenzato dalla natura selvatica di *T. urartu*, in natura infatti le piante devono essere in grado di svilupparsi il più velocemente possibile in modo da aumentare la loro *chance* di sopravvivenza in un'ambiente già densamente popolato. La differenza riscontrata nello sviluppo delle piante potrebbe anche essere influenzata dal portamento, che risulta infatti essere differente tra le accessioni (Figura 29 a). Un campione con portamento prostrato può dimostrare dimensioni maggiori rispetto a una pianta con portamento eretto, e questo solo da un punto di vista visivo. Quando il seme raccolto durante questa tesi sarà nuovamente seminato, tutto il materiale avrà condiviso il medesimo *iter* di conservazione. Piantando un eguale numero di semi, sarà possibile quantificare l'effettiva capacità di crescita ed accestimento tra le diverse accessioni.

La natura selvatica di *T. urartu* è stata immediatamente visibile in tutta la collezione anche grazie a due caratteri di rusticità evidenti, il nodo pubescente e il colore violaceo della foglia bandiera (Figura 24b-c). I tricomi sono uno strumento spesso utilizzato dalle piante per difendersi da eventuali attacchi, la pubescenza infatti rende più difficile l'adesione dell'insetto alla pianta stessa. La perdita di questi caratteri selvatici nelle linee coltivate, è la probabile conseguenza della domesticazione e della dipendenza dall'uomo (Hawkes 2013; Zohary, Hopf, e Weiss 2012; Brown *et al.* 2009). Oltre a questi caratteri subito evidenti, è stata riscontrata la presenza di possibili alleli di resistenza per *Puccinia striiformis* f. sp. *tritici* (ruggine gialla). Durante la prova in campo, infatti, si sono sviluppati diversi focolai di infezione, probabilmente contribuiti da un'estesa infezione di ruggine gialla nei campi adiacenti. Nelle aree infette internamente al nostro

campo sperimentale erano presenti piante evidentemente malate e, adiacenti a queste, accessioni totalmente sane. In alcuni casi queste erano intercalate a più accessioni infettate. La resistenza di queste accessioni può essere contribuita da alleli di resistenza a questa patologia fungina, come già dimostrato in altri frumenti selvatici (Uauy *et al.* 2005). Partendo dall'evidente manifestazione fenotipica di questi caratteri selvatici, è plausibile formulare l'ipotesi della possibile presenza di ulteriori alleli di rusticità che però non presentano un fenotipo evidente, come quelli coinvolti all'adattamento ambientale. Studi mirati alla scoperta di questi alleli permetteranno di caratterizzarne l'importanza nell'ottica del miglioramento genetico di frumento. In seguito alla scoperta di questi, si potrà procedere al trasferimento dei caratteri più utili, alle varietà attualmente coltivate di frumento attraverso tecniche tradizionali (*crossing* e *backcrossing*) o biotecnologiche (*Targeting* delle sequenza omologhe o trasformazione genica).

La distribuzione normale degli altri caratteri fenotipici (Figura 28, 29, 30) suggerisce che la collezione raccoglie un gradiente di variabilità poco strutturato per tutti i caratteri analizzati. Questo è in accordo con il risultato emerso dai dati molecolari (Figura 22), ove la varianza dei due assi di PC si limita a valori inferiori al 50%. Entrambi questi risultati dimostrano come la collezione analizzata fornisca una buona rappresentazione della variabilità di *T. urartu*. Questa caratteristica rende quindi la collezione uno strumento promettente per l'identificazione delle basi genetiche dei caratteri complessi e adattativi con tecniche statistiche derivate dal *QTL mapping*.

Nonostante la maggior parte delle accessioni presentasse per i caratteri fenotipici analizzati valori nella media, le code nei grafici di distribuzione indicavano valori estremi. Questa diversità si è evidenziata in misura maggiore in alcuni caratteri come per esempio

l'altezza delle piante (Figura 28a) e l'epoca di spigatura (Figura 28b) in cui, per l'appunto, la distribuzione dei campioni risultava essere molto ampia. Considerando l'altezza, erano presenti fenotipi bassi, cioè con altezze inferiori 100 cm, e accessioni estremamente alte, fino a 199 cm. La diversità esistente si è evidenziata maggiormente con l'epoca di fioritura, la cui distribuzione presentava un *range* straordinariamente ampio: la spigatura è iniziata alla fine di Aprile con le accessioni più precoci, e si è conclusa a fine Maggio con quelle più tardive. Questo arco temporale di 3 settimane è un'ulteriore indice della notevole variabilità presente all'interno della collezione.

La diversità fenotipica manifesta e finanche inattesa di alcune accessioni (Figura 27), ci ha spinto ad un'indagine più approfondita circa la provenienza degli stessi dalle *seedbank* che hanno fornito il germoplasma. In particolare, la banca dati NIAS, pur avendo distribuito accessioni etichettate come *T. urartu*, ha avanzato la possibilità che alcune di esse potessero essere fatte risalire a *T. timopheevii*. Questa specie tetraploide, originata dall'ibridazione di *T. urartu* e *A. speltoides*, potrebbe essere stata inclusa nella collezione ed essere responsabile di parte degli *outlier* geografici, fenotipici e molecolari riscontrati. Questo può essere dovuto al fatto che, popolazioni naturali di *T. timopheevii* e *T. urartu* crescono ancora oggi nelle medesime regione dell'Armenia, è dunque facile ipotizzare che sia avvenuta una contaminazione all'epoca del campionamento.

Con lo scopo di affinare la caratterizzazione fenotipica, gli studi futuri saranno condotti con un numero maggiore di semi e con repliche per ciascuna accessione. In questo modo, sarà possibile rilevare i caratteri agronomici in modo più solido e fornire le basi per analisi di genetica quantitativa volte alla scoperta di loci responsabili per i

caratteri di interesse agronomico che la collezione di *T. urartu* ha già mostrato in questa tesi.

6 Conclusioni

L'elevata variabilità della collezione è stata confermata a tutti e 3 i livelli di analisi preposti. La caratterizzazione climatica fornisce informazioni significative e non ridondanti, i dettagli unici ottenuti dalle variabili bioclimatiche sottendono infatti alla probabilità della presenza di alleli adattativi tra le accessioni.

Il notevole numero di SNP prodotti sul *subset* di 24 campioni è indice dell'elevata variabilità allelica esistente. Le analisi molecolari hanno evidenziato l'inconsistenza tra la distanza geografica e genetica: mentre alcune accessioni si raggruppano in *cluster* monofiletici in base al sito di campionamento, gli altri campioni presentano similarità a livello molecolare andando a costituire gruppi pur avendo origini geograficamente distanti. Questo è una dimostrazione che la distanza geografica non è il solo, o il principale, fattore che contribuisce al differenziamento molecolare. Nonostante l'esiguo numero di campioni analizzati con la tecnica di RADSeq, la bassa strutturazione emersa indica la presenza di un gradiente di variabilità e quindi di una buona rappresentazione della collezione di *T. urartu* assemblata.

Di centrale importanza è stata la caratterizzazione fenotipica di tutta la collezione condotta in campo. Grazie ai rilievi agronomici effettuati durante la stagione di crescita di *T. urartu* è stato possibile dimostrare l'ampia diversità presente all'interno della collezione. La distribuzione normale di questi caratteri fenotipici risulta essere imprescindibile per studiare le basi genetiche dei caratteri complessi di *T. urartu* e porre le fondamenta per future analisi di *QTL mapping*.

7 Ringraziamenti

Come prevedibile anche io sono caduta nel *clichè* dei ringraziamenti, ma in fondo che tesi sarebbe senza di essi? A questo proposito le prime persone che mi sento in dovere di ringraziare sono il Professore Mario Enrico Pè, ed il Prof. Rodolfo Bernardi, i miei due Relatori: senza la loro fiducia e la loro guida sapiente non avrei potuto compiere questo percorso formativo che ha arricchito il mio bagaglio personale con nuove esperienze, conoscenze e amicizie.

Desidero ringraziare anche Matteo dell'Acqua, Co-relatore, che mi ha aiutato nella stesura della tesi con suggerimenti, critiche ed osservazioni: a lui va la mia gratitudine, anche se a me spetta la responsabilità per ogni errore contenuto in questa tesi.

Ringrazio anche i colleghi o meglio amici dell'università che hanno reso questi due anni indimenticabili, in particolare Alice Pieri e Alice Brunazzi perché sono loro che mi hanno sopportato e supportato in tutti questi giorni.

Infine vorrei ringraziare le persone a me più care: i mie genitori, perché mi hanno sempre sostenuto in tutte le decisioni e il mio fidanzato Jacopo, perché in questi anni pieni di cambiamenti e nuovi inizi lui è stato il mio punto fermo, la persona che mi è sempre stata vicina anche nei momenti peggiori e che mi ha aiutato a superare le difficoltà.

Ringrazio ancora tutte le persone, anche quelle che per motivi di spazio non ho nominato, che fanno parte della mia vita semplicemente perché solo facendone parte contribuiscono a renderla meravigliosa.

8 Bibliografia

- Ambaglio D., Foraboschi D. Le civiltà dell'antichità 1, Milano, Edizioni Scolastiche Mondadori, 1994.
- Bar-Yosef O. 1998. "The Natufian Culture in the Levant, Threshold to the Origins of Agriculture." *Evolutionary Anthropology: Issues, News, and Reviews* 6 (5): 159–77. doi:10.1002/(SICI)1520-6505(1998).
- Becker R. A., Chambers J. M., and Wilks A. R. 1988. *The New S Language: A Programming Environment for Data Analysis and Graphics*. Monterey, CA, USA: Wadsworth and Brooks/Cole Advanced Books & Software.
- Bradbury P. J., Zhang Z., Kroon D. E., Casstevens T. M., Ramdoss Y., and Buckler E. S. 2007. "TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples." *Bioinformatics (Oxford, England)* 23 (19): 2633–35. doi:10.1093/bioinformatics/btm308.
- Brown T. A. 1999. "How Ancient DNA May Help in Understanding the Origin and Spread of Agriculture." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 354 (1379): 89–98. doi:10.1098/rstb.1999.0362.
- Brown T. A., Jones M. K., Powell W., and Allaby R. G. 2009. "The Complex Origins of Domesticated Crops in the Fertile Crescent." *Trends in Ecology & Evolution* 24 (2): 103–9. doi:10.1016/j.tree.2008.09.008.
- Catchen J., Hohenlohe P. A., Bassham S., Amores A., and Cresko W. A. 2013. "Sacks: An Analysis Tool Set for Population Genomics." *Molecular Ecology* 22 (11): 3124–40. doi:10.1111/mec.12354.
- Clayton W.D., Vorontsova M.S., Harman K.T. and Williamson H. (2006 onwards). GrassBase - The Online World Grass Flora. <http://www.kew.org/data/grasses-db.html>.
- Crop Wild Relatives and Climate Change* (2013) Online resource. Accessed on 11-12-2014. www.cwrdiversity.org
- Davey J. W. and Blaxter M. L. 2010. "RADSeq: Next-Generation Population Genetics." *Briefings in Functional Genomics* 9 (5-6): 416–23. doi:10.1093/bfgp/elq031.
- Dell'Acqua M., Zuccolo A., Tuna M., Gianfranceschi L., and Pè M. E. 2014. "Targeting Environmental Adaptation in the Monocot Model Brachypodium Distachyon: A

Multi-Faceted Approach.” *BMC Genomics* 15 (1): 801. doi:10.1186/1471-2164-15-801.

Diamond J. 2015. “Armi, Acciaio E Malattie. Breve Storia Del Mondo Negli Ultimi Tredicimila Anni.” *Www.libreriauniversitaria.it*. Accessed August 20. <http://www.libreriauniversitaria.it/armi-acciaio-malattie-breve-storia/libro/9788806183547>.

Dvorak J. and Akhunov E. D. 2005. “Tempos of Gene Locus Deletions and Duplications and Their Relationship to Recombination Rate During Diploid and Polyploid Evolution in the Aegilops-Triticum Alliance.” *Genetics* 171 (1): 323–32. doi:10.1534/genetics.105.041632.

Elshire R. J., Glaubitz J. C., Sun Q., Poland J. A., Kawamoto K., Buckler E. S., and Mitchell S. E. 2011. “A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species.” *PLoS ONE* 6 (5): e19379. doi:10.1371/journal.pone.0019379.

Feldman M., Lupton F.G.H., Miller T.E. (1995) Wheats. *Triticum* spp. (Gramineae-Triticinae). In: Smartt J, Simmonds NW (eds) *Evolution of crop plants*. Longman Scientific & Technical Press, London, pp 84–192.

Feldman M. and Levy A.A. 2005. “Allopolyploidy – a Shaping Force in the Evolution of Wheat Genomes.” *Cytogenetic and Genome Research* 109 (1-3): 250–58. doi:10.1159/000082407.

Feuillet C., Langridge P. and Waugh R. 2008. “Cereal breeding takes a walk on the wild side”. *Trends in Genetics*. 24 (1), 24–32. doi:10.1016/j.tig.2007.11.001.

Fricano A., Brandolini A., Rossini L., Sourdille P., Wunder J., Effgen S., Hidalgo A., Erba D., Piffanelli P., and Salamini F. 2014. “Crossability of *Triticum Urartu* and *Triticum Monococcum* Wheats, Homoeologous Recombination, and Description of a Panel of Interspecific Introgression Lines.” *G3 (Bethesda, Md.)* 4 (10): 1931–41. doi:10.1534/g3.114.013623.

Fu D., Uauy C., Distelfeld A., Blechl A., Epstein L., Chen X., Sela H., Fahima T., and Dubcovsky J. 2009. “A Kinase-START Gene Confers Temperature-Dependent Resistance to Wheat Stripe Rust.” *Science* 323 (5919): 1357–60. doi:10.1126/science.1166289.

Fuller D. Q. 2007. “Contrasting Patterns in Crop Domestication and Domestication Rates: Recent Archaeobotanical Insights from the Old World.” *Annals of Botany* 100 (5): 903–24. doi:10.1093/aob/mcm048.

- Giles R. J. and Brown T. A. 2006. "GluDy Allele Variations in *Aegilops Tauschii* and *Triticum Aestivum*: Implications for the Origins of Hexaploid Wheats." *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 112 (8): 1563–72. doi:10.1007/s00122-006-0259-5.
- Gill B. S., Friebe B., Raupp W. J., Wilson D. L., Cox T. S., Sears R. G., Brown-Guedira G. L., and Fritz A. K. 2006. "Wheat Genetics Resource Center: The First 25 Years." In *Advances in Agronomy*, edited by Donald L. Sparks, 89:73–136. Academic Press. <http://www.sciencedirect.com/science/article/pii/S0065211305890029>.
- Gorham J., Bristol A., Young E. M., and Wyn J. R. G. 1991. "The Presence of the Enhanced K/Na Discrimination Trait in Diploid *Triticum* Species." *Theoretical and Applied Genetics* 82 (6): 729–236. doi:10.1007/BF00227318.
- Gupta P. K., Balyan H. S., Sharma P. C., and Ramesh B. 1996. "Microsatellites in Plants: A New Class of Molecular Markers." *Current Science* 45 (1): 45–54.
- Gustafson P., Raskina O., Ma X., and Nevo E. 2009a. "Wheat Evolution, Domestication, and Improvement." In *Wheat Science and Trade*, edited by Brett F. Carver Editor Regentessor, 3–30. Wiley-Blackwell. <http://onlinelibrary.wiley.com/doi/10.1002/9780813818832.ch1/summary>.
- Hammami R., Jouve N., Soler C., Frieiro E., and González J. M. 2014. "Genetic Diversity of SSR and ISSR Markers in Wild Populations of *Brachypodium Distachyon* and Its Close Relatives *B. Stacei* and *B. Hybridum* (Poaceae)." *Plant Systematics and Evolution* 300 (9): 2029–40. doi:10.1007/s00606-014-1021-0.
- Harlan J. R., de Wet J. M. J., and Glen Price E.. 1973. "Comparative Evolution of Cereals." *Evolution* 27 (2): 311–25. doi:10.2307/2406971.
- Haudry A., Cenci A., Ravel C., Bataillon T., Brunel D., Poncet C., Hochu I. 2007. "Grinding up Wheat: A Massive Loss of Nucleotide Diversity Since Domestication." *Molecular Biology and Evolution* 24 (7): 1506–17. doi:10.1093/molbev/msm077.
- Hawkes J. G. 2013. *The Diversity of Crop Plants*. Harvard University Press.
- Hijmans R. J., Cameron S. E., Parra J. L., Jones P. G., and Jarvis A. 2005. "Very High Resolution Interpolated Climate Surfaces for Global Land Areas." *International Journal of Climatology* 25 (15): 1965–78. doi:10.1002/joc.1276.

- Hillman G. C. and Davies M. S. 1990. "6. Domestication Rates in Wild-Type Wheats and Barley under Primitive Cultivation." *Biological Journal of the Linnean Society* 39 (1): 39–78. doi:10.1111/j.1095-8312.1990.tb01611.x.
- Hoyt E. 1988. *Conserving the Wild Relatives of Crops*, Rome, Italy, IPGRI/IUCN/WWF
- Huang S., Sirikhachornkit A., Su X., Faris J., Gill B., Haselkorn R. and Gornicki P. 2002. "Genes Encoding Plastid Acetyl-CoA Carboxylase and 3-Phosphoglycerate Kinase of the Triticum/Aegilops Complex and the Evolutionary History of Polyploid Wheat." *Proceedings of the National Academy of Sciences of the United States of America* 99 (12): 8133–38. doi:10.1073/pnas.072223799.
- Jackson M., Ford-Lloyd B. and Parry M. 2013a. *Plant Genetic Resources and Climate Change*. CABI.
- Janick J. and Gepts P. 2010. *Plant Breeding Reviews, Part 2: Long-Term Selection: Crops, Animals, and Bacteria*. John Wiley & Sons.
- Kara Y., Martín A., Souyris I., Rekika D., and Monneveux P. 2000. "Root Characteristics in Durum Wheat (*T. Turgidum* Conv. *Durum*) and Some Wild Triticeae Species. Genetic Variation and Relationship with Plant Architecture." *Cereal Research Communications* 28 (3): 247–54.
- Kilian B., Ozkan H., Deusch O., Effgen S., Brandolini A., Kohl J., Martin W. and Salamini F. 2007. "Independent Wheat B and G Genome Origins in Outcrossing *Aegilops* Progenitor Haplotypes." *Molecular Biology and Evolution* 24 (1): 217–27. doi:10.1093/molbev/msl151.
- Kriticos D. J., Webber B. L., Leriche A., Ota N., Macadam I., Bathols J., and Scott J. K. 2012. "CliMond: Global High-Resolution Historical and Future Scenario Climate Surfaces for Bioclimatic Modelling." *Methods in Ecology and Evolution* 3 (1): 53–64. doi:10.1111/j.2041-210X.2011.00134.x.
- Ling H.-Q., Zhao S., Liu D., Wang J., Sun H., Zhang C., Fan H. 2013. "Draft Genome of the Wheat A-Genome Progenitor *Triticum Urartu*." *Nature* 496 (7443): 87–90. doi:10.1038/nature11997.
- Liu X. M., Brown-Guedira G. L., Hatchett J. H., Owuochi J. O. and Chen M.-S. 2005. "Genetic Characterization and Molecular Mapping of a Hessian Fly-Resistance Gene Transferred from *T. Turgidum* Ssp. *Dicoccum* to Common Wheat." *Theoretical and Applied Genetics* 111 (7): 1308–15. doi:10.1007/s00122-005-0059-3.

- Liverani M. 2009. “Antico Oriente.” *Goodreads*.
http://www.goodreads.com/work/best_book/15377556-antico-oriente-storia-societ-economia.
- Luikart G., England P. R., Tallmon D., Jordan S. and Taberlet P. 2003. “The Power and Promise of Population Genomics: From Genotyping to Genome Typing.” *Nature Reviews Genetics* 4 (12): 981–94. doi:10.1038/nrg1226.
- Ma J., Zhou R., Dong Y., Wang L., Wang X., and Jia J. 2001. “Molecular Mapping and Detection of the Yellow Rust Resistance Gene Yr26 in Wheat Transferred from *Triticum Turgidum* L. Using Microsatellite Markers.” *Euphytica* 120 (2): 219–26. doi:10.1023/A:1017510331721.
- Manel S., Joost S., Epperson B. K., Holderegger R., Storer A., Rosenberg M. S., Scribner K. T., Bonin A. and Fortin M.-J. 2010. “Perspectives on the Use of Landscape Genetics to Detect Genetic Adaptive Variation in the Field.” *Molecular Ecology* 19 (17): 3760–72. doi:10.1111/j.1365-294X.2010.04717.x.
- Mardia K. V., Kent J. T. and Bibby J. M. 1979. *Multivariate Analysis*. Academic Press.
- Matsuoka Y.. 2011. “Evolution of Polyploid Triticum Wheats under Cultivation: The Role of Domestication, Natural Hybridization and Allopolyploid Speciation in Their Diversification.” *Plant and Cell Physiology* 52 (5): 750–64. doi:10.1093/pcp/pcr018.
- Matsuoka Y. and Nasuda S. 2004. “Durum Wheat as a Candidate for the Unknown Female Progenitor of Bread Wheat: An Empirical Study with a Highly Fertile F1 Hybrid with *Aegilops Tauschii* Coss.” *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 109 (8): 1710–17. doi:10.1007/s00122-004-1806-6.
- Moghaddam M., Ehdaie B. and Waines J. G. 2000. “Genetic Diversity in Populations of Wild Diploid Wheat *Triticum Urartu* Tum. Ex. Gandil. Revealed by Isozyme Markers.” *Genetic Resources and Crop Evolution* 47 (3): 323–34. doi:10.1023/A:1008782532505.
- Peng J. H., Sun D. and Nevo E. 2011a. “Domestication Evolution, Genetics and Genomics in Wheat.” *Molecular Breeding* 28 (3): 281–301. doi:10.1007/s11032-011-9608-4.
- Pirez V. I., de Carvalho F. I. F., Benin G., da Silveira G., da Silva J. A. G., Nornberg R., Hagemann T., de Souza Luche H. and de Oliveira A. C. 2013. “Seeding Density

- in Wheat: The More, the Merrier?" *Scientia Agricola* 70 (3): 176–84. doi:10.1590/S0103-90162013000300006.
- Plucknett D., Smith N., Williams J., Murthi Anishetty N. 1987. *Gene Banks and the World's Food*. Princeton University Press, Princeton, NJ.
- Qiu Y. C., Zhou R. H., Kong X. Y., Zhang S. S. and Jia J. Z. 2005. "Microsatellite Mapping of a Triticum Urartu Tum. Derived Powdery Mildew Resistance Gene Transferred to Common Wheat (Triticum Aestivum L.)." *Theoretical and Applied Genetics* 111 (8): 1524–31. doi:10.1007/s00122-005-0081-5.
- R Development Core Team. 2015. "R: A Language and Environment for Statistical Computing." Text. *GBIF.ORG*. <http://www.gbif.org/resource/81287>.
- Reynolds M., Dreccer F. and Trethowan R. 2007. "Drought-Adaptive Traits Derived from Wheat Wild Relatives and Landraces." *Journal of Experimental Botany* 58 (2): 177–86. doi:10.1093/jxb/erl250.
- Rouse M. N. and Jin Y. 2011. "Stem Rust Resistance in A-Genome Diploid Relatives of Wheat." *Plant Disease* 95 (8): 941–44. doi:10.1094/PDIS-04-10-0260.
- Salamini F., Özkan H., Brandolini A., Schäfer-Pregl R., and Martin W. 2002. "Genetics and Geography of Wild Cereal Domestication in the near East." *Nature Reviews Genetics* 3 (6): 429–41. doi:10.1038/nrg817.
- Schlötterer C. and Pemberton J. 1998. "The Use of Microsatellites for Genetic Analysis of Natural Populations — a Critical Review." In *Molecular Approaches to Ecology and Evolution*, edited by Dr Rob DeSalle and Dr Bernd Schierwater, 71–86. Birkhäuser Basel. http://link.springer.com/chapter/10.1007/978-3-0348-8948-3_4.
- Storfer A., Murphy M. A., Spear S. F., Holderegger R. and Waits L. P. 2010. "Landscape Genetics: Where Are We Now?" *Molecular Ecology* 19 (17): 3496–3514. doi:10.1111/j.1365-294X.2010.04691.x.
- Tanksley S. D. and McCouch S. R. 1997. "Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild." *Science (New York, N.Y.)* 277 (5329): 1063–66.
- Turner M. G., Gardner R. H. and O'Neill R.V. 2001. *Landscape Ecology in Theory and Practice: Pattern and Process*. Springer Science & Business Media.
- Uauy C., Brevis J. C., X. Chen, Khan I., Jackson L., Chicaiza O., Distelfeld A., Fahima T. and Dubcovsky J. 2005. "High-Temperature Adult-Plant (HTAP) Stripe Rust

- Resistance Gene Yr36 from *Triticum Turgidum* Ssp. *Dicoccoides* Is Closely Linked to the Grain Protein Content Locus Gpc-B1.” *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 112 (1): 97–105. doi:10.1007/s00122-005-0109-x.
- Vadez V., Hash T., Bidinger F. and Kholova J. 2012. “II.1.5 Phenotyping Pearl Millet for Adaptation to Drought.” *Plant Physiology* 3: 386. doi:10.3389/fphys.2012.00386.
- Valkoun, J. J. 2001. “Wheat Pre-Breeding Using Wild Progenitors.” *Euphytica* 119 (1-2): 17–23. doi:10.1023/A:1017562909881.
- Venables W. N. and Ripley B. D. 2002. *Modern Applied Statistics with S. Statistics and Computing*. New York, NY: Springer New York. <http://link.springer.com/10.1007/978-0-387-21706-2>.
- Vos P., Hogers R., Bleeker M., Reijans M., van de Lee T., Hornes M., Friters A. 1995. “AFLP: A New Technique for DNA Fingerprinting.” *Nucleic Acids Research* 23 (21): 4407–14. doi:10.1093/nar/23.21.4407.
- Wei T. 2013. *Corrplot: Visualization of a Correlation Matrix* (version 0.73). <https://cran.r-project.org/web/packages/corrplot/index.html>.
- Wright S. 1943. “Isolation by Distance.” *Genetics* 28 (2): 114–38.
- Zohary D. and 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*. Oxford University Press.
- Zohary D., Hopf M. and Weiss E. 2012. *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin*. OUP Oxford.