



UNIVERSITÀ DI PISA

Dipartimento di Scienze Agrarie Alimentari e Agro- ambientali

Laurea magistrale in Biotecnologie vegetali e microbiche

**Genomica comparata di isolati di
Trichoderma agenti di biocontrollo**

Candidata

Giulia Piaggese

Correlatore

Prof. Mario Enrico Pè

Relatori

Prof. Giovanni Vannacci
Dott.ssa Sabrina Sarrocco

Anno Accademico 2013-2014

*“Fare ciò che ami è libertà,
amare ciò che fai è felicità.”*
(*Anonimo*)

INDICE

Abstract	5
1. INTRODUZIONE	6
1.1 Trichoderma spp.	6
1.1.1 Generalità e morfologia.....	6
1.1.2 <i>Trichoderma</i> come agente di biocontrollo e microrganismo benefico.....	8
1.1.3 Isolati fungini di <i>Trichoderma</i> oggetto di studio del presente lavoro.....	14
1.1.3.1 <i>Trichoderma harzianum</i> T6776.....	14
1.1.3.2 <i>Trichoderma gamsii</i> T6085.....	15
1.2 Genomica fungina	16
1.2.1 Time lapse della definizione di “genomica”.....	16
1.2.2 Perché sequenziare i genomi fungini.....	17
1.2.3 Storia e risorse.....	18
1.2.4 Tecniche di sequenziamento dei genomi fungini.....	20
1.3 Genomica comparata	23
1.3.1 General overview.....	24
1.3.2 Espansioni geniche in relazione alla specializzazione ecologica.....	25
1.3.3 Famiglie geniche espanse nei micoparassiti e agenti di biocontrollo.....	26
1.3.3.1 CaZY (<i>Carbohydrate Active Enzymes</i>).....	28
1.3.3.2 Peptidasi.....	29
1.3.3.3 Classi proteiche coinvolte nella sintesi di metaboliti secondari.....	30
2. SCOPO DEL LAVORO	31
3. MATERIALI E METODI	32

3.1 Isolati fungini.....	32
3.2 Estrazione del DNA genomico.....	32
3.3 Sequenziamento.....	34
3.3.1 Valutazione del sequenziamento.....	34
3.4 Assemblaggio del genoma.....	35
3.4.1 Valutazione della completezza dell'assemblaggio.....	36
3.5 Predizione e annotazione genica.....	36
3.6 Comparazione genomica.....	38
3.6.1 Scelta degli organismi.....	38
3.6.2 Analisi comparata.....	38
3.7 Analisi del secretoma.....	41
4. RISULTI E DISCUSSIONE.....	42
4.1 Valutazione della qualità del DNA genomico.....	42
4.2 Valutazione del sequenziamento.....	42
4.3 Valutazione dell'assemblaggio.....	51
4.4 Predizione genica.....	54
4.5 Analisi di comparazione genomica.....	55
4.6 Famiglie geniche espanse in <i>Trichoderma</i>	57
4.6.1 Cell-wall-degrading enzymes (CWDEs).....	58
4.6.2 Peptidasi.....	60
4.6.3 Proteine coinvolte nella sintesi di metaboliti secondari.....	61
4.6.4 Famiglie proteiche non appartenenti a classi specifiche.....	62
4.7 Caratterizzazione dei CaZY (<i>Carbohydrate Active Enzymes</i>).....	67
5. CONCLUSIONI.....	70
6. BIBLIOGRAFIA.....	72

Abstract

Molti isolati di *Trichoderma* sono ben noti per i loro effetti benefici che si esplicano sia in termini di protezione delle piante contro fattori biotici e abiotici che di promozione della crescita. L'isolato *Trichoderma gamsii* 6085 è da tempo noto come agente di lotta biologica, in condizioni di laboratorio e in campo, contro gli agenti causali del *Fusarium* Head Blight (FHB) o Fusariosi del frumento, in particolare *Fusarium graminearum* e *Fusarium culmorum*, di cui, grazie alla sua abilità di crescere in presenza di deossinivalenolo, ne limita la crescita e la produzione di micotossine.

Trichoderma harzianum 6776 svolge la sua azione benefica stimolando la crescita di pomodoro, attraverso un effetto di stimolo dell'attività fotosintetica, produzione di ormoni e presenza di zuccheri nella pianta. Inoltre è in grado di indurre resistenza a stress biotici ed abiotici.

Nella presente tesi il DNA totale genomico di entrambi gli isolati è stato sequenziato mediante la tecnologia MiSeq Illumina, con sequenze mate pair-ends di 250bp. Le sequenze ottenute (copertura media 80X per *T. harzianum* e 90X per *T. gamsii*) sono state assemblate con Velvet versione 1.2.07.

Il genoma assemblato di *T. gamsii* 6085 è risultato essere composto da 381 scaffolds con una lunghezza totale di 37.97 Mbp mentre quello di *T. harzianum* 6776 da 1573 scaffolds con una lunghezza totale di 39.73 Mbp. La completezza del genoma è stata valutata utilizzando CEGMA, che ha stimato la sequenza genomica essere completa al 98.39 % per *T. harzianum* 6776 e 87,58% per *T. gamsii* 6085). I genomi sono stati annotati in modo automatico con la pipeline MAKER2. Complessivamente sono stati identificati 10.944 modelli genici codificanti proteine contenute nel genoma nucleare in *T. gamsii* e 11.501 in *T. harzianum*. I proteomi predetti dei genomi degli isolati sequenziati sono stati comparati con altre specie appartenenti al genere *Trichoderma* e con un insieme di organismi aventi comportamento biologico simile e differente, in modo da poter individuare le espansioni delle famiglie geniche associate alle interazioni di agente di biocontrollo e di micoparassitismo degli isolati oggetto di studio.

In dettaglio sono state caratterizzate le espansioni delle famiglie geniche delle classi enzimatiche appartenenti ai CaZY (*Carbohydrate-Active enZymes*) coinvolte nel processo di micoparassitismo di *Trichoderma*.

Tale studio ha l'obiettivo di fornire una piattaforma utile per ulteriori studi futuri di genomica funzionale quali analisi di trascrittomica, metabolomica e proteomica.

1.Introduzione

1.1 *Trichoderma spp.*

1.1.1 Generalità e morfologia

Il genere *Trichoderma* fu identificato per la prima volta nel 1794 da Persoon. Vi appartengono specie con una forte adattabilità ecologica, capaci di crescere su substrati molto differenti tra loro (terreno, legno, cortecce, tessuti interni delle piante ospiti). Sono specie pleomorfe, la cui forma anamorfa è propria del genere *Trichoderma*, mentre quella telomorfa appartiene al genere *Hypocrea* (ordine Hypocreales, classe Pyrenomycetes, divisione Ascomycota) (Druzhinina *et al.*, 2011). La relazione tra le due forme, anamorfa e teleomorfa, fu verificata soltanto nel 1957 per opera di Dingley e in seguito, da Webster e Rifai (Rifai and Webster, 1996; Webster and Rifai, 1968).

Gli individui appartenenti al genere *Trichoderma* generalmente possiedono elevate velocità di accrescimento; le colonie sono caratterizzate da un micelio aereo ialino che può assumere un aspetto variabile tra disordinato, fioccoso, lanoso o aracnoide in base al ceppo ed al mezzo di coltura. Alcuni isolati possono modificare il colore del substrato di coltura rilasciando pigmenti di colore giallo, ambra, verde, rosso tenue o marrone chiaro. Una delle caratteristiche peculiari di questo genere è la produzione di composti che conferiscono al micelio di alcune specie un caratteristico odore di cocco o canfora, con intensità variabile tra il pronunciato ed il fortissimo.

Le colonie di *Trichoderma* fatte crescere su piastre di PDA (*Potato Dextrose Agar*) si presentano, spesso, a cerchi concentrici, con alternanza di colorazione della colonia verde / incolore, dovuta all'alternarsi di zone ricche di spore con zone dove la sporulazione è assente. Quest'alternanza è dovuta a ritmi circadiani che simulano l'alternanza giorno/notte (Fig.1.1).

La classificazione e l'identificazione delle specie di *Trichoderma* a livello morfologico è estremamente difficile a causa delle differenze spesso minime dal punto di vista fenotipico tra diverse specie (Samuels, 2006; Rifai, 1969; Bissett *et al.*, 2003). L'avanzamento delle tecniche di biologia molecolare ha permesso di poter identificare le specie analizzando le sequenze di rDNA 18S e dell'ITS (Internal Transcribed Spacer), specifiche per ogni specie (Druzhinina e Kopchinskiy, 2006; Hagan *et al.*, 2007). Questo contributo ha fatto sì che nel

corso degli anni siano state classificate un numero sempre maggiore di specie, fino ad arrivare ad oggi con più di 100 specie identificate in Europa (Field and Druzhinina, 2012).

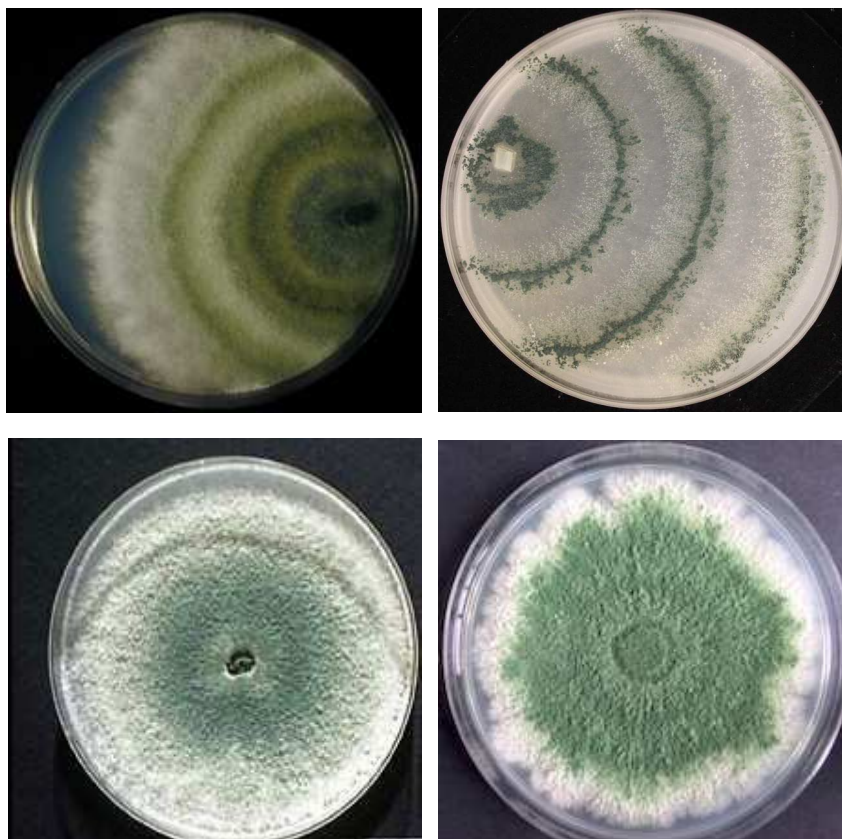


Figura. 1.1 Colonie di isolati di *Trichoderma* spp. su PDA; in alto a sinistra *T. virens*; a destra *T. longibrachiatum*; in basso a sinistra *T. harzianum*; in basso a destra *Trichoderma* spp.

Al genere *Trichoderma* appartengono funghi ubiquitari a rapida crescita, estremamente comuni nei suoli agricoli, nelle foreste, nelle praterie ed in suoli desertici di tutte le zone climatiche. Si tratta di funghi saprofiti, prolifici produttori di proteine extracellulari, tra le quali un vasto assortimento di enzimi litici come cellulasi e chitinasi (Klein e Eveleigh, 1998) e più di 100 differenti metaboliti ad attività antibiotica nota (Kubicek and Harman, 1998; Sivasithamparam and Ghisalberti, 1998). Questo ampio spettro enzimatico conferisce loro la capacità di trasformare un'ampia varietà di materiale sia organico sia xenobiotico (Druzhinina *et al.*, 2012).

Le specie appartenenti al genere *Trichoderma* possiedono un metabolismo estremamente versatile: possono utilizzare come fonte di carbonio un'ampia varietà di substrati, inoltre sono uno dei pochi gruppi di organismi capaci di utilizzare composti costituiti da un solo

atomo di carbonio. Molti ceppi riescono a degradare gli idrocarburi e questo è dimostrato anche dal fatto che costituiscono la componente maggiore della popolazione fungina presente in suoli contaminati. Come fonte di azoto sono capaci di utilizzare dai composti ammoniacali agli amminoacidi, alle proteine, mentre l'assimilazione di nitrato è scarsa e specie specifica. Gli isolati hanno la capacità di mobilizzare e assimilare i nutrienti presenti nel terreno, tali da competere con altri microorganismi tellurici (Benitez *et al.*, 2004). È stata riscontrata una tolleranza a diversi composti usualmente tossici nei confronti di organismi fungini, quali fungicidi (Harman *et al.*, 1996), fitoalessine di origine vegetale (Howell *et al.*, 2000) ed alcune tossine anche di origine fungina (Matarese *et al.*, 2012). La base molecolare di questo meccanismo di resistenza è stata parzialmente delucidata con la scoperta che l'espressione di trasportatori ABC (ATP binding cassette) in *Trichoderma* è indotta dalla presenza di funghi fitopatogeni o dai loro metaboliti (Lanzuise *et al.*, 2002).

L'abbondanza di *Trichoderma* spp. in vari suoli è da ascrivere all'ampia versatilità metabolica, alla capacità di colonizzare dinamicamente la rizosfera della pianta, all'abilità di competere e reprimere un gran numero di patogeni nonché alla facilità di adattamento alle diverse situazioni ambientali ed alle ridotte necessità nutrizionali. Una delle caratteristiche più interessanti è che *Trichoderma* è in grado di colonizzare l'apparato radicale della pianta, ma non risulta patogeno dal momento che, tranne pochissime eccezioni, limita la colonizzazione alle cellule superficiali delle radici impedendo la crescita o la penetrazione di altri patogeni (Elad *et al.*, 1999; Elad e Kapat, 1999; Yesdida *et al.*, 1999). Nonostante *Trichoderma* abbia una capacità intrinseca di attaccare le piante, è solitamente avirulento, avendo sviluppato una relazione con le piante di tipo simbiotico più che parassitario (Harman *et al.*, 2004).

L'insieme di tutte le caratteristiche qui elencate fa sì che isolati appartenenti al genere *Trichoderma* siano spesso considerati agenti di biocontrollo (BCA) assumendo un ruolo importante all'interno di strategie di lotta biologica contro le malattie delle piante. Infatti, il genere *Trichoderma* include quasi, se non oltre, il 50% di specie fungine utilizzate nella lotta biologica (Verma, 2007).

1.1.2 *Trichoderma* come agente di biocontrollo e microrganismo benefico

La sfida maggiore alla quale saremo chiamati a rispondere nell'immediato futuro è quella di soddisfare la richiesta di cibo per una popolazione mondiale in continuo accrescimento con una sempre più pressante necessità di produrre alimenti che non

contengano residui chimici e che prevedano l'utilizzo di strategie di coltivazione e difesa ecosostenibili. In questa ottica il controllo biologico si inserisce perfettamente anche all'interno di un sistema di lotta integrata, rispondendo al decreto legislativo n. 150 del 14 agosto 2012, che recepisce la direttiva comunitaria 2009/128/CE in cui vengono definite le misure per un uso sostenibile dei prodotti fitosanitari al fine di ridurre i rischi sulla salute umana, sull'ambiente e sulla biodiversità, promuovendo l'applicazione di metodi non chimici.

Con il termine "agente di biocontrollo" si indica un microrganismo benefico in grado di contrastare l'attacco di organismi fitopatogeni e, nella più moderna definizione, esercitare un effetto benefico sulla pianta ospite (Shoresh *et al.*, 2010). Al genere *Trichoderma* appartengono numerose specie, per lo più di origine tellurica, ampiamente utilizzate come agenti di biocontrollo contro patogeni di grande importanza agricola (Chet and Inbar 1994; Harman *et al.* 2004; Hermosa *et al.*, 2012).

L'azione di biocontrollo messa in atto da *Trichoderma* spp. è il risultato di diversi meccanismi:

- **Micoparassitismo:** si tratta di un fenomeno legato ad una successione di eventi quali riconoscimento, attacco, penetrazione e morte dell'ospite fungino. Durante questo processo i funghi parassitizzano e conducono a morte i loro ospiti attraverso la formazione di avvolgimenti ifali (*coilings*, Fig.1.2), la secrezione di enzimi litici e la penetrazione diretta dell'ospite. Il meccanismo attrattivo nei confronti dell'ospite potrebbe essere indotto dal rilascio di lectine da parte della preda (Druzhinina *et al.*, 2011). *Trichoderma* secreta un complesso di enzimi litici (di cui le chitinasi sono il gruppo principale) che permettono di degradare la parete cellulare dell'ospite consentendo il rilascio di sostanze nutritive e oligomeri (Kubicek *et al.*, 2001; Howell 2003; Woo *et al.*, 2006; Vinale 2008) da cui il micoparassita trae fonte nutritiva. Dal punto di vista evolutivo, il micoparassitismo si ritiene essere un carattere ancestrale diffuso all'interno del genere *Trichoderma*, con *T. virens* e *T. atroviride* considerate le specie che meglio hanno sviluppato questo tipo di stile di vita. Si pensa che il micoparassitismo abbia in qualche modo anche facilitato lo sviluppo di tutte quelle interazioni definite "positive" fra *Trichoderma* e le piante (Druzhinina *et al.*, 2011).

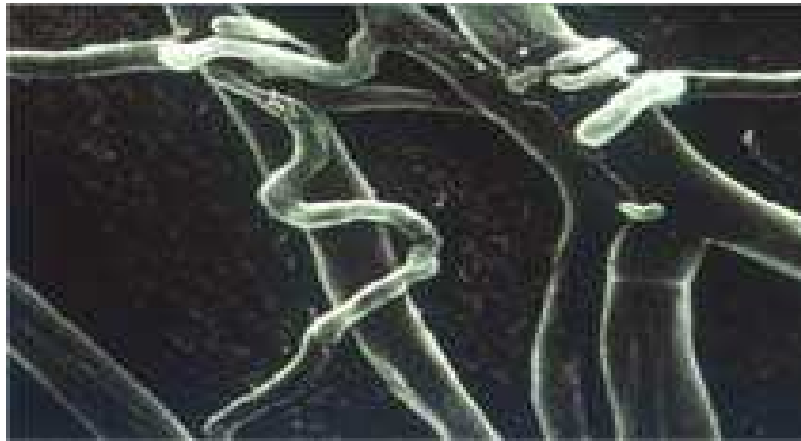


Figura.1.2. Coilings di *T. asperellum* intorno alle ife di *R. solani*

- **Antibiosi:** consiste nella produzione di una vasta gamma di metaboliti secondari come acidi organici, peptaboliti e composti eterociclici. Si tratta di sostanze, volatili o diffusibili, tossiche per il patogeno in grado di inibirne o bloccarne la crescita. La loro efficacia è stata dimostrata più volte come ad esempio in un lavoro di Ghisalberti e altri (2004) in cui è stato dimostrato il contenimento di *Gaeumannomyces graminis* var. *tritici* da parte di *T. harzianum* per opera di furani.

Nei filtrati colturali di *T. harzianum* (ceppi T22 e T39) sono stati ritrovati diversi metaboliti prodotti durante l'antagonismo verso *R. solani*. Sono state identificate due vie principali: la produzione di composti a basso peso molecolare, non polari, volatili che si trovano in alte concentrazioni nel terreno con effetto a lunga durata sui microorganismi; oppure la produzione di composti polari e peptaboliti che agiscono rapidamente sulle β -glucanasi bloccando la produzione di ife (Schirmbock *et al.*, 1994; Lorito *et al.*, 1996; Fogliano *et al.*, 2002, Vinale *et al.*, 2008).

- **Competizione:** si tratta di un meccanismo solitamente messo in atto in un ambiente circoscritto nel quale due o più organismi hanno bisogno della stessa risorsa, la cui disponibilità ed utilizzo sono limitati. Si può verificare competizione per le sostanze nutritive (carbonio e azoto), per lo spazio di crescita e per i siti d'infezione. Ad esempio, *T. harzianum* riesce a controllare l'attacco da parte di *B. cinerea* sui tessuti dell'acino di uva occupando il sito d'infezione (Gullino, 1992). Silvan e Chet (1989) hanno dimostrato che la competizione per i nutrienti è il

maggior meccanismo utilizzato da *T. harzianum* per controllare *F. oxysporum* f. sp. *melonis*.

Fino a pochi anni fa questi venivano considerati i principali meccanismi di biocontrollo. Studi più recenti hanno dimostrato che l'azione benefica che svolge *Trichoderma* spp. è legata principalmente all'interazione instaurata tra pianta e fungo il quale, attraverso la colonizzazione radicale, provoca l'induzione di cambiamenti a livello metabolico e cellulare migliorando la resistenza di quest'ultima ai patogeni. Questa ipotesi si contrappone al vecchio modello che vedeva come azione principale quella diretta tra *Trichoderma* e il patogeno nella difesa della pianta (Fig.1.3).

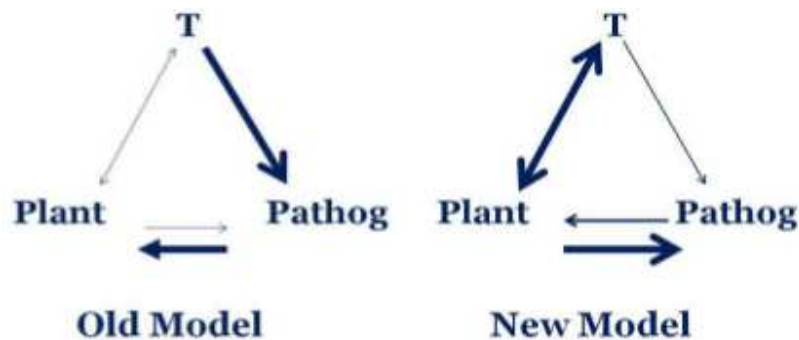


Figura 1.3. Interazione *Trichoderma*-pianta-patogeno (Harman *et al.*, 2004)

L'interazione *Trichoderma*-pianta implica il riconoscimento a livello molecolare attraverso una rete di segnali mediati da ormoni prodotti dalla pianta come: l'acido salicilico (SA), l'acido jasmonico (JA) e etilene (ET). Questi ormoni sono noti come molecole coinvolte nella via della trasduzione del segnale per la resistenza sistemica indotta (ISR) nella pianta data dalla presenza di *Trichoderma* all'interno degli organi vegetali. Il segnale di trasduzione prende inizio con l'accumulo di acido salicilico (SA), proprio come nel pathway della resistenza sistemica acquisita (SAR) indotta dall'attacco di patogeni. Questa azione si configura come "effetto priming" nello scatenare le risposte della pianta che si concretizzano in una resistenza nei confronti di fattori di stress biotici e abiotici e in una migliore efficienza dell'assimilazione dell'azoto (NUE) (Shoresh and Harman, 2010; Hermosa *et al.*, 2012) (Fig.1.4).

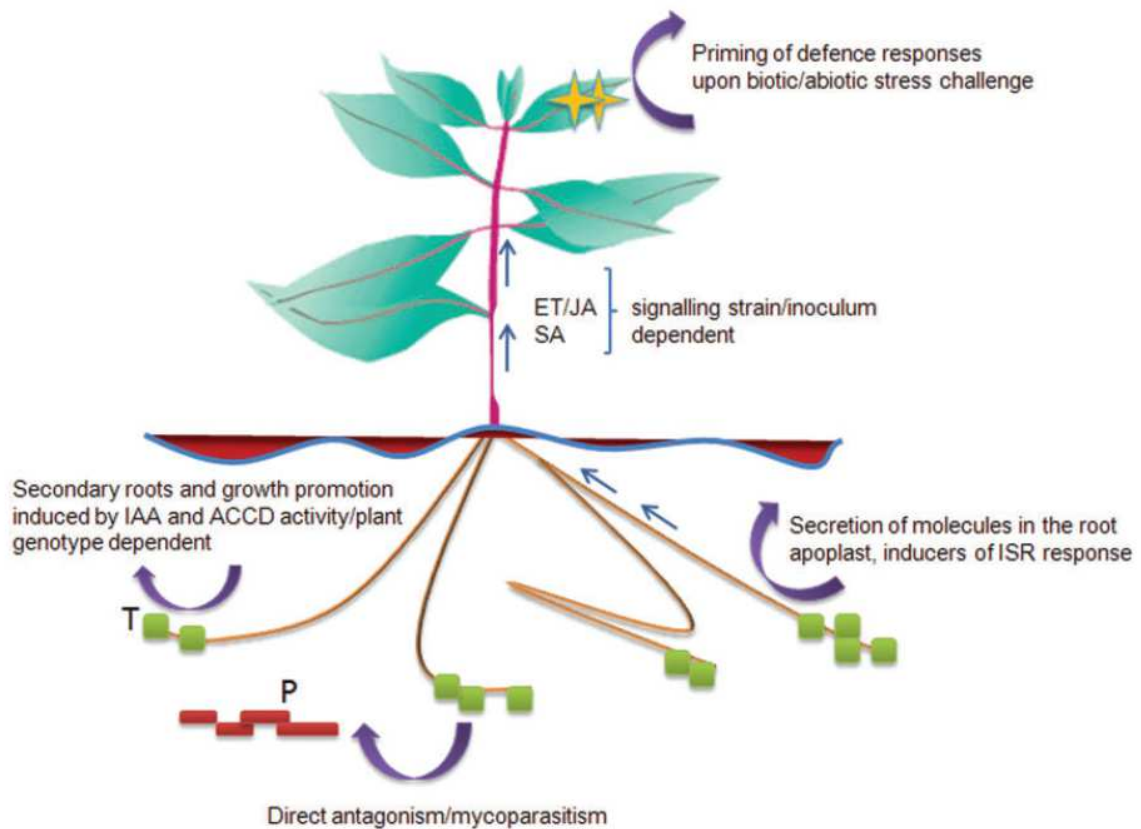


Figura 1.4. Meccanismi di risposta della pianta stimolati dalla colonizzazione radicale da parte di *Trichoderma* (Hermosa *et al.*, 2012)

Oltre che stimolare le risposte di difesa della pianta, *Trichoderma* incrementa il processo fotosintetico, migliora il rendimento della produttività, favorisce la crescita sia dell'apparato radicale che di quello aereo. Inoltre, la crescita dell'apparato radicale indotta da *Trichoderma* spp. aumenta l'assorbimento dei nutrienti, la tolleranza alla siccità del suolo, favorisce la germinazione e la vigoria dei semi (Harman *et al* 2004; Shoresh *et al* 2010). In Figura 1.5 sono riportati i contributi scientifici più significativi della storia di *Trichoderma*.

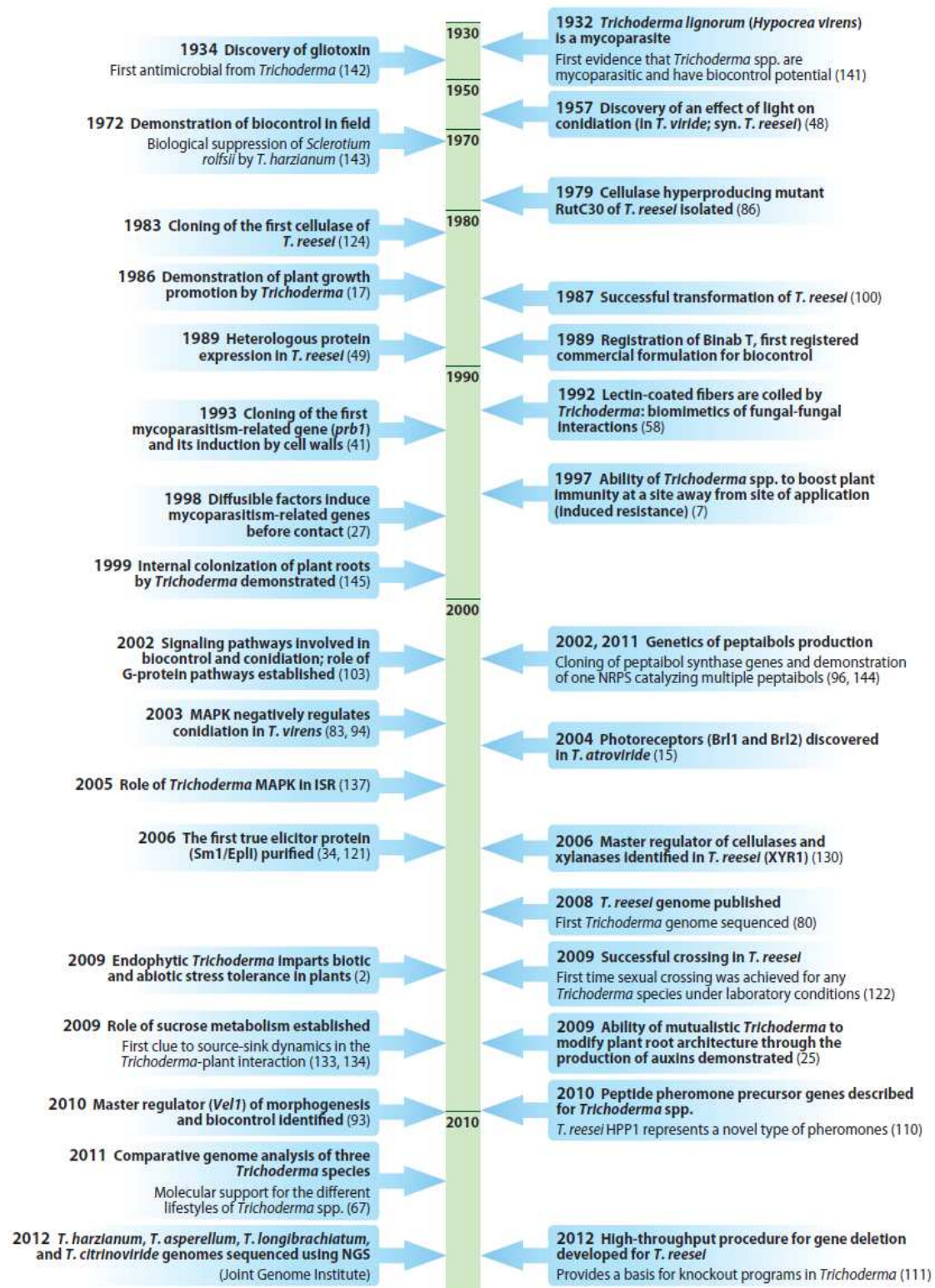


Figura 1.5. Contributi delle scoperte scientifiche di maggior rilevanza sui meccanismi di azione di *Trichoderma* (Mukherjee *et al.*, 2013)

1.1.3 Isolati fungini di Trichoderma oggetto di studio del presente lavoro

1.1.3.1 Trichoderma harzianum T6776

T. harzianum T6776, isolato per la prima volta in Italia dal terreno di un pescheto a Colignola (Pisa), è noto per la sua attività come agente di biocontrollo, micoparassita e promotore della crescita di molte specie vegetali coltivate (Sarrocco *et al.*, 2013; Moncini *et al.*, 2013). Queste attività sono connesse alla capacità di questo organismo di crescere come endofita nelle radici delle piante senza causare danni diretti all'ospite.

Sono stati condotti alcuni studi riguardanti l'azione di biocontrollo di *T. harzianum* T6776 nei confronti di patogeni tellurici, tra cui *Rhizoctonia solani*, *Fusarium oxysporum* f. sp. *lycopersici* e *Fusarium oxysporum* f. sp. *radicis-lycopersici*, sia in ravanello che in diverse cultivar di pomodoro mediante diverse procedure di inoculazione, mostrando un possibile impiego come biofungicida (Sarrocco *et al.*, 2013; Fiorini *et al.*, 2014).

In generale è stato mostrato che *T. harzianum* T6776 svolge la sua azione come endofita promuovendo la crescita delle colture che si traduce in un incremento dell'altezza dello stelo e sviluppo della lamina fogliare ed un aumento della velocità di germinazione, e riducendo il ciclo produttivo della specie vegetale con conseguenze favorevoli in termini di abbattimento dei costi di produzione (Moncini *et al.*, 2013). Studi recenti dimostrano che l'influenza positiva di T6776 come endofita in piante di pomodoro (Micro-Tom) cresciute in coltura idroponica sia dovuta ad un miglioramento dell'efficienza fotosintetica e traslocazione degli zuccheri all'interno degli organi vegetali ed ad un miglior bilanciamento ormonale che potrebbero aumentare la risposta della pianta a stress di natura biotica e abiotica (Fiorini *et al.*, 2014)(Fig.1.6).

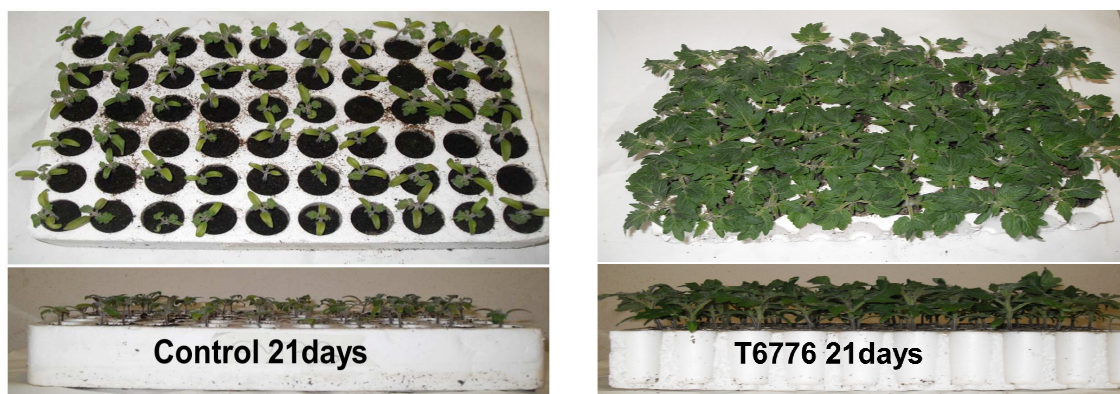


Figura 1.6. Effetti benefici di *T. harzianum* T6776 in piantine di pomodoro (Micro-Tom) cresciute in coltura idroponica. Nell'immagine sulla sinistra sono presenti le plantule senza essere state inoculate con T6776, nell'immagine a fianco viene mostrato l'incremento di altezza, vigoria e sviluppo fogliare delle plantule inoculate con l'isolato T6776.

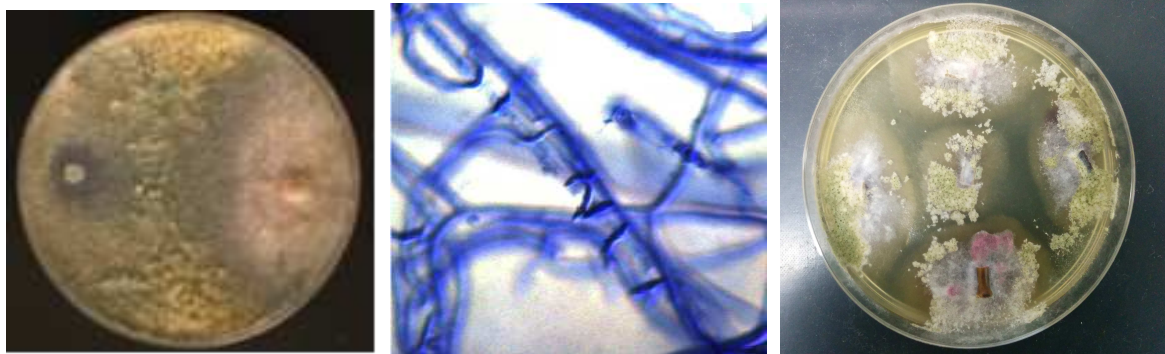
1.1.3.2 *Trichoderma gamsii* T6085

L'isolato *T. gamsii* T6085, proveniente da un terreno non coltivato in Crimea, Ucraina, è stato ampiamente studiato per le sue capacità di biocontrollo della fusariosi della spiga di frumento, Fusarium head blight (FHB), grazie alla sua capacità di ridurre la crescita di *Fusarium culmorum* e *Fusarium graminearum* in pieno campo (Sarrocco *et al.*, 2013), di crescere in presenza di elevate dosi della micotossina deossinivalenolo (DON) riducendone la produzione da parte dei patogeni sopra citati, e di inibire la crescita attraverso diversi meccanismi d'azione (Matarese *et al.*, 2012). Non sono ancora chiari i meccanismi coinvolti nella resistenza al DON ma l'ipotesi più verosimile, in attesa di conferma, vedrebbe il coinvolgimento dei trasportatori ABC e dei MFS (Major Facilitator Superfamily) (Matarese *et al.*, 2012; Sarrocco *et al.*, 2013), mentre è stata esclusa la sua capacità di degradare la micotossina o di modificarne la struttura. Sono, inoltre, a disposizione alcuni studi d'interazione molecolare volti a capire i meccanismi coinvolti nel processo di micoparassitismo tra l'isolato T6085 verso le due specie di *Fusarium* mostrando il coinvolgimento di numerose chitinasi (Matarese *et al.*, 2012).

Attualmente sono in corso studi che vedono il coinvolgimento di *T. gamsii* T6085 come agente di biocontrollo nei confronti di *F. graminearum* e *F. culmorum*, agenti causali della fusariosi della spiga del frumento (FHB) e produttori di micotossine DON e NIV, all'interno di un approccio di multitrofico, in abbinamento ad isolati di *F. oxysporum*. Dalle prime prove di competizione per substrati di crescita, condotti *in vitro*, risulta un effetto di contenimento della crescita di *F. graminearum* dovuto alla presenza di *T. gamsii* T6085 e di

F. oxysporum, sia da soli che in combinazione (Valenti, dati non pubblicati).

Questa mancata interferenza tra i due organismi apre scenari di notevole importanza al fine di sviluppare strategie di lotta biologica che vedono la co-presenza dei due antagonisti in modo da garantire un'efficacia maggiore nel controllo dei patogeni.



A

B

C

Figura 1.7. Micoparassitismo *T. gamsii* T6085: a) inibizione della crescita di *F. graminearum* da parte di *T. gamsii* T6085; b) avvolgimenti ifali (coilings) di *T. gamsii* T6085 intorno a *F. culmorum*; c) inibizione della crescita di *F. graminearum* e di *F. oxysporum* da parte di *T. gamsii* T6085.

1.2 Genomica fungina

1.2.1 Time lapse della definizione di “genomica”

La genomica può essere definita come l'evoluzione della genetica classica.

Il termine “genomica” è stato coniato per la prima volta nel 1986 da Thomas Roderick per descrivere il susseguirsi delle azioni di mappatura, sequenziamento ed analisi di genomi (Heiter and Boguski., 1997)

In passato, quando i citogenetisti parlavano di “genoma”, si riferivano ad un singolo set completo di cromosomi che venivano osservati al microscopio per visualizzare il cariotipo. Il genoma veniva utilizzato per studi citologici per l'identificazione del progenitore più antico (antenato) che prendeva parte alla formazione di una specie esistente, oppure veniva impiegato in studi per determinare l'omologia tra set cromosomici osservando la migrazione di quest'ultimi durante le fasi meiotiche.

Oggi, il senso moderno della parola “genoma” sta ad indicare l'insieme di sequenze di DNA genomico ottenute mediante tecniche di sequenziamento automatizzate in larga scala

e spesso assemblate in lunghe porzioni di DNA ma, non sempre, in cromosomi finiti (Bennet and Arnold., 2001). L'obiettivo principale del sequenziamento del genoma è quello di riuscire a predire ed annotare il set completo di geni presenti in un organismo: questo permette non solo di determinare la struttura, la funzione e le relazioni evolutive dei geni all'interno di organismi della stessa specie o tra quelli di specie differenti, ma anche di poter svolgere studi di genetica funzionale, trascrittomica e metabolomica in modo da rivelare le interazioni tra l'organismo e l'ambiente in cui vive.

1.2.2 Perché sequenziare i genomi fungini

L'ultimo decennio ha visto una rivoluzione della genomica del regno fungino. Il sequenziamento del primo genoma fungino risale a quello di *Saccharomyces cerevisiae* nel 1996 (Goffeau *et al.* 1996); ad oggi il numero di sequenze genomiche disponibili è aumentato di quasi oltre un ordine di grandezza.

Il quinto regno annovera oltre 1,5 milioni di membri che hanno un impatto e un'interazione sia come “amico” che come “nemico”, nei confronti di molteplici substrati e forme di vita presenti nel pianeta (Galagan *et al.*, 2005).

I funghi svolgono ruoli di primaria importanza all'interno della catena trofica ambientale agendo sia da decompositori di materiale organico e non, sia attraverso rapporti ecologici di tipo parassitario, simbiotico e/o mutualistico con procarioti, piante (comprese le alghe) ed animali.

Sono organismi con una forte capacità adattativa, in grado di sopravvivere in ambienti a pH estremo, subire e resistere a forti stress ambientali conservandosi sotto forma di strutture di sopravvivenza (come spore, sclerozi, ecc.). Queste caratteristiche sono dovute ad un metabolismo estremamente versatile, in grado di produrre una vasta gamma di acidi, enzimi litici, proteine a basso peso molecolare e numerosi metaboliti secondari, che ne assicurano la nutrizione, la sopravvivenza, la colonizzazione ed il mantenimento del possesso di specifiche nicchie ecologiche.

Nella tassonomia moderna il regno fungino risulta separato da quello vegetale e animale (Bennett J. W., 1997). I funghi condividono una lunga storia con la civiltà umana: i primi riferimenti sono stati trovati all'interno della letteratura greca, dove si parla di “pietre fungine” risalenti alla civiltà Maya (del 1000-300 aC) (Galagan *et al.*, 2005) e molti dei più antichi processi biotecnologici si basano sull'utilizzo dei funghi come la fermentazione o produzione di antibiotici. Col passare del tempo le grandi industrie hanno sviluppato un

notevole interesse economico per questi organismi, utilizzandoli per la produzione commerciale di enzimi litici come glucoamilasi, pectinasi, proteasi, lipasi e cellulasi, impiegate nella trasformazione alimentare come edulcoranti, nella preparazione di succhi di frutta, pane e formaggi (Bennet., 2007).

Nel corso degli anni i numerosi studi di fisiologia cellulare e genetica fungina hanno dimostrato la presenza di caratteri che rendono stretta la correlazione o vicinanza di questi ultimi con il regno vegetale e animale: pluricellularità, presenza del citoscheletro, sviluppo e differenziazione sessuale, riproduzione, ciclo cellulare, meccanismi di segnalazione intercellulare, influenza dei ritmi circadiani, metilazione del DNA e modifiche della cromatina. I geni responsabili di tutte queste funzioni biologiche hanno origini comuni, tanto da fare sì che gli studi sulla genetica fungina siano di primaria importanza anche per lo sviluppo degli studi nel campo umano e vegetale (Galagan *et al.*, 2005). Sono inoltre ritenuti uno dei sistemi modello più utili in biologia cellulare per la loro facilità di accrescimento e gestibilità in fase sperimentale ma anche per la presenza di un genoma molto piccolo che ne facilita, in termini di qualità e tempo impiegato, il sequenziamento, l'assemblamento e l'annotazione.

1.2.3 Storia e risorse

Nonostante l'importanza e l'utilità di questi organismi, fino a poco tempo fa tutto ciò che era conosciuto sui genomi fungini derivava dal genoma del lievito *S. cerevisiae* (Goffeau *et al.* 1996). Questo lavoro ha rivoluzionato completamente le conoscenze sui lieviti ed ha costituito una base di partenza per gli studi di funzione, espressione e interazione dei geni eucariotici (Galagan *et al.*, 2005; Grigoriev., 2013). Tuttavia la sequenza del genoma di lievito ha fornito solo un assaggio limitato della biodiversità del regno fungino. La vera rivoluzione è avvenuta più recentemente, prima con il sequenziamento del genoma di *Schizosaccharomyces pombe*, il quale ha completato le conoscenze sulla natura del gene eucariotico, permettendo il passaggio dalla presenza di un unico esone in *S. cerevisiae* ad una visione multi- esonica (Wood *et al.*, 2002; Grigoriev., 2013) e quindi con quello del primo fungo filamentoso *Neurospora crassa* (Galagan *et al.*, 2003).

La disponibilità della sequenza genomica di *N. crassa* ha rivelato i limiti e la diversità esistenti tra i lieviti e i funghi filamentosi, mostrando una taglia del genoma e un numero di geni doppi rispetto a *S. cerevisiae*. Oltre a queste caratteristiche era presente anche un elevato numero di sequenze ripetute e introni. Il genoma di lievito scarseggiava di questi

due elementi e mancava di oltre il 40% delle proteine omologhe a geni presenti in *N. crassa* (Galagan *et al.*, 2005). Queste scoperte hanno dato il via a numerosi studi evolutivi che hanno permesso di comprendere i meccanismi di duplicazione genica dei genomi dei funghi filamentosi e quelli legati alla perdita dei geni nei lieviti associati ai processi di speciazione di entrambi questi organismi (Dujon., 2006).

Nonostante evidenti necessità, il progresso del sequenziamento dei genomi fungini inizialmente era lento. Per accelerare il ritmo della genomica fungina nel 2000, un consorzio di micologi in collaborazione con gli scienziati del Broad Institute hanno lanciato il Fungal Genome Initiative (FGI; Cuomo and Birren., 2010) il cui obiettivo era quello di sequenziare i genomi di funghi provenienti da tutto il regno (Galagan *et al.*, 2005; Grigoriev., 2013).

L'aumento del numero dei genomi sequenziati è stato accompagnato da un numero pressoché uguale di progetti e collaborazioni tra i più grandi centri mondiali di ricerca come: il Joint Genome Institute (JGI) che ha lanciato nel 2009 il Fungal Genomics Program (<http://igi.doe.gov/fungi>) focalizzandosi sui funghi importanti per l'energia e l'ambiente (Grigoriev., 2013), il Washington University Genome Sequencing Center, il TIGR, il Sanger Institute, il Marine Biological Laboratories (MBL), lo Stanford Genome Technology Center, il Duke Center for Genome Research, l'University of British e generati attraverso il supporto di numerose agenzie di finanziamento tra cui il National Human Genome Research Institute (NHGRI), il National Science Foundation ed i Dipartimenti dell'Agricoltura e dell'Energia degli Stati Uniti (Galagan *et al.*, 2005). In particolare, è stata osservata anche una crescente collaborazione tra il mondo accademico e quello industriale, che ha visto il coinvolgimento delle più note multinazionali come Monsanto, Syngenta e Bayer.

Nel 2005 più di 40 sequenze di genomi fungini erano a disposizione del pubblico con oltre 40 ulteriori progetti in corso (Galagan *et al.*, 2005): oggi ne sono disponibili più di 200 (Grigoriev., 2013).

Questi genomi appartengono ad importanti agenti patogeni umani, patogeni delle piante, saprofiti e organismi modello. È importante sottolineare che la maggior parte dei genomi disponibili sono divisi in raggruppamenti che vanno a rappresentare le principali divisioni fungine (Ascomycota, Basidiomycota, Zigomycota, e Chytridiomycota), in modo da consentire analisi comparate al fine di valutarne le distanze evolutive ed i tratti legati alla loro ecologia.

I dati riguardanti i genomi fungini sequenziati sono pubblici e disponibili all'interno di

diversi database online presenti nei siti Web del Broad Institute - nella sezione Fungal Genome Initiative (<http://www.broad.mit.edu/FGI/>), nel database MycoCosm del JGI (<http://jgi.doe.gov/fungi>), nel database fungino del TIGR (www.tigr.org), nella banca dati di NCBI (<http://www.ncbi.nlm.nih.gov/>) e molti altri ancora.

1.2.4 Tecniche di sequenziamento dei genomi fungini

La rivoluzione nel campo della genomica fungina è stata guidata dall'evoluzione delle tecnologie di sequenziamento del genoma (Galagan *et al.*, 2005).

I primi sequenziamenti si basavano sull'applicazione del metodo basato sulla biochimica di Sanger che ha dominato le scene dagli anni '70 fino ai primi anni del 2000, passando dal metodo classico (detto anche enzimatico) descritto da Sanger e Gilbert nel 1977 fino a quello semi-automatico che ha sostituito la corsa elettroforetica originale con un' elettroforesi di tipo capillare, aumentando la velocità di separazione dei frammenti di DNA sul gel e l'efficienza del sequenziamento. Gli aspetti innovativi risiedevano nell'uso del nuovo biopolimero poliacrilammide nel gel e nel poter analizzare in parallelo più campioni per corsa. Nel corso degli anni sono state messe in atto numerose modifiche a questo metodo fino ad arrivare nel 2000 al moderno sequenziatore Sanger (Applied Biosystems, ABI 3730x1) a 96 capillari, capace di produrre sequenze di circa 1000bp, per un totale di 96kb in 3 ore di corsa. Nonostante la lunghezza delle sequenze e l'accuratezza al 99,9% dovuta al basso numero di errori a carico delle basi, l'applicazione di questo metodo nei grandi progetti di sequenziamento è sempre stata ridotta a causa dei costi eccessivi.

I primi genomi fungini, e la maggior parte di essi, sono stati ottenuti con il metodo Sanger seguendo l'approccio whole genome shotgun (WGS) che prevedeva la costruzione di una libreria genomica contenete inserti di varie dimensioni, sequenziando solo le estremità degli inserti dei vari cloni e assemblati in contig (Shendure and Ji., 2008). Questo approccio rappresenta un progresso rispetto a quello clone-by-clone, metodo molto laborioso e lungo che prevedeva il clonaggio, sequenziamento e assemblamento per ogni clone generato (Grigoriev., 2013). L'approccio WGS ha notevolmente accelerato i tempi di sequenziamento riducendone drasticamente i costi. Il primo genoma sequenziato con questo approccio è stato quello del batteriofago di 48.5kbp, che ha fatto da apri fila all'era dei sequenziamenti di interi genomi come quello del lievito *S. cerevisiae* nel 1996 (Goffeau, Barrell *et al.*, 1996), seguito, poco dopo da altri due ascomiceti,

Schizosaccharomyces pombe (Wood *et al.*, 2002) e *Neurospora crassa* (Galagan *et al.*, 2003) e dal genoma di *Phanerochaete chrysosporium* (Martinez *et al.*, 2004).

Il sequenziamento Sanger ha dominato la genomica per più di due decenni, fino alla nascita delle tecnologie di nuova generazione (NGS) che prevedono la produzione di un elevato numero di nucleotidi per corsa. Si tratta di tecniche ad elevato parallelismo, dove non è più necessaria la migrazione dei frammenti su gel ma si parla di letture in tempo reale sintetizzando i frammenti di DNA. Queste metodiche di nuova concezione riducono i tempi ed i costi di sequenziamento, a discapito dell'accuratezza e lunghezza delle sequenze, producendo frammenti inferiori a 500bp. Inoltre, prevedono l'amplificazione dei frammenti di DNA tramite delle reazioni di PCR eliminando la fase di clonaggio (Shendure and Ji., 2008). Le tecnologie NGS si diversificano tra loro per la chimica utilizzata, la lunghezza delle sequenze prodotte (bp), il contenuto in GC, la quantità e precisione dei dati prodotti. Tra le tecnologie disponibili in commercio, la macchina Illumina Solexa HiSeq 2000 è in grado di produrre una mole di dati che possono arrivare fino a mezzo Tb in un'unica corsa, con sequenze lunghe fino a 150 bp. La sua "sorella minore", MiSeq, invece può produrre sequenze fino a 2×250 bp in tempi estremamente ridotti ("overnight run"). La metodica che prevede il pirosequenziamento di La Roche, 454, produce sequenze più lunghe (fino a 1 Kbp con l'ultima versione della macchina FLX) dell'Illumina, ma ad un costo più elevato e con un numero maggiore di errori dovuti al sequenziamento.

Uno degli ultimi arrivi fra le tecniche NGS è la Pacific Biosciences (PacBio): si tratta di macchine che producono sequenze molto lunghe fino a vari Kbp, con tasso di errore frequente (fino al 15%), ma distribuito in modo casuale in tutta la lunghezza della reads, e può essere applicato utilizzando un'unica molecola DNA per il sequenziamento di un intero genoma. Infine, Oxford Nanopore Technologies ha annunciato una nuova macchina che offre un collegamento USB al proprio pc dove è possibile sequenziare un intero genoma a partire da singola molecola di DNA (Pennisi, 2012).

	Run Time	Read Length	Quality	Total nucleotides sequenced	Cost /MB
454 Pyrosequencing	24h	700 bp	Q20-Q30	0.7 GB	\$10
Illumina Miseq	27h	2x250bp	> Q30	15 GB	\$0.15
Illumina Hiseq 2500	11days	2x125bp	>Q30	1000 GB	\$0.05
Ion torrent	2h	400bp	>Q20	50MB-1GB	\$1
Pacific Biosciences	2h	10-20kb	>Q30 consensus >Q10 single	400-800MB /SMRT cell	\$0.33-\$1

Figura 1.8. Caratteristiche delle principali tecnologie NGS in commercio.

Questi approcci, ognuno con i propri punti di forza e debolezze, possono essere anche combinati tra loro per ottenere assemblamenti di elevata qualità, poiché la produzione di brevi sequenze causa delle difficoltà di assemblamento richiedendo elevati valori di copertura (Morozova and Marra., 2008). Molti genomi sono stati prodotti con approcci ibridi combinando dati provenienti da tecniche di sequenziamento NGS come 454, Illumina ma anche con sequenze prodotte con il metodo Sanger. Ad esempio, nel progetto di sequenziamento delle specie di *Colletotrichum* sono stati combinati i dati provenienti da sistemi Illumina con quelli del 454 (O'Connell *et al.*, 2012).

L'approccio ibrido viene impiegato quando sono necessari elevati valori di copertura (dai 30X ai 100X) per assemblare genomi *de novo* utilizzando le "short reads" che, oltre ad essere molto brevi, con molta probabilità potrebbero contenere al loro interno errori. La combinazione con le sequenze Sanger che oltre ad avere dimensioni maggiori (di 3-10 kb) richiedono anche coperture meno elevate (dai 6X ai 10X), permette di avere un minor tasso di errori delle basi.

Queste tecniche vengono utilizzate anche in progetti di risequenziamento e finitura dei genomi, incrementando il livello di copertura del genoma, l'espansione dei contigs, il processo di scaffolding, riducendo il numero di gaps, migliorando la qualità del sequenziamento e aumentando la qualità della chiamata delle basi.

Il genoma di *S. cerevisiae* non solo è il primo genoma fungino ad essere stato sequenziato ma è stato anche il primo ad essere finito. Ci sono voluti 13 anni prima che i primi genomi

di funghi filamentosi fossero portati a termine, nonostante il gran numero di sequenze quasi complete come in *N. crassa* (Galagan *et al.*, 2003). I fattori che concorrono alla presenza di genomi non finiti sono legati alla difficoltà di assemblare le sequenze ripetute, la presenza di polimorfismi e di geni duplicati, oltre ai gap fisici dovuti alla mancata copertura del genoma a causa della tecnologia di sequenziamento scelta. Un genoma completo offre molte più informazioni di uno non finito: corretto ordine di geni, presenza o assenza di un gene e distinzioni tra geni e pseudogeni. I geni ripetuti nelle regioni subtelomeriche possono essere assemblati in modo non corretto o omessi nei progetti di assemblamento, ma allo stesso tempo possono portare importanti informazioni biologiche, spesso legate alle esigenze ecologiche e specializzazioni, come ad esempio i geni di virulenza di agenti patogeni (Grigoriev., 2013). Sebbene la maggior parte del contenuto genico all'interno di un genoma di 40 Mb potrebbe essere rivelato attraverso una corsa Illumina, la sua completezza è incerta fino a quando il genoma non è finito e accuratamente annotato.

Se con il metodo Sanger la rifinitura di un genoma richiede un costo proibitivo, le tecniche NGS offrono nuove possibilità di miglioramento del genoma, come l'utilizzo di una combinazione delle tecnologie Illumina con quella PacBio. Anche se lunghe sequenze PacBio hanno un tasso di errore del 15%, questi errori possono essere corretti mediante le elevate coperture ottenute tramite i dati Illumina; inoltre le lunghe sequenze possono servire da riferimento per l'assemblaggio in contig delle short reads Illumina eliminando eventuali gap. Una copertura tra il 30X e il 50X delle sequenze PacBio sono sufficienti per assemblare in modo finito le sequenze genomiche fungine (Grigoriev., 2013).

Queste tecnologie hanno trovato molteplici campi di applicazioni che vanno dal sequenziamento del DNA per studi di genomica comparata, genomica funzionale, metagenomica e epigenomica al sequenziamento dell'RNA per studi riguardanti i profili di espressione come i micro-RNA e la trascrittomica (RNA-seq) (Shendure and Ji, 2008; Morozova and Marra., 2008; Grigoriev, 2013).

1.3 Genomica comparata

“La genomica comparata ha lo scopo di individuare i tratti genici riguardanti i processi evolutivi legati alla storia e ai cambiamenti della sequenza genomica.

La grande varietà di stili di vita, tipi di riproduzione e metabolismo presenti nel regno fungino, fanno sì che questi organismi siano un modello perfetto per studiare i processi di adattamento in funzione delle modifiche dell'architettura genomica” (Leducq., 2014)

1.3.1 General overview

Il sequenziamento dei genomi produce un'elevata mole di dati che richiedono un'analisi attenta e mirata ai fini di individuare caratteristiche specifiche di un organismo.

La genomica comparata rappresenta un potente strumento di analisi dei dati genomici mettendo a confronto due o più sequenze per far luce sulle diversità degli organismi presi in esame.

Alla base della suddetta analisi vi sta il contenuto genico presente in un genoma. Principalmente viene utilizzata per individuare le famiglie geniche espanse o perse all'interno di un organismo, migliorando anche il processo di annotazione dei geni. Annotare i genomi degli eucarioti è molto impegnativo perché richiede la combinazione di più approcci ma soprattutto non vi sono modelli generali sviluppati come per i genomi batterici, se annotiamo più genomi contemporaneamente utilizzando anche quelli appartenenti ad organismi modello, come *N. crassa* e *M. grisea*, l'inesattezza della predizione genica tende a diminuire e si avvalora (Grigoriev., 2013).

L'analisi di comparazione genomica è stata utilizzata fin dal sequenziamento dei primi organismi appartenenti a questo regno, il primo fungo filamentoso sequenziato, *N. crassa* fu confrontato con quello di *S. cerevisiae* rilevando le differenze tra organismi unicellulari e quelli pluricellulari (Dujon.,2004).

Nel corso degli anni si sono succeduti una serie di lavori di comparazione genomica e di tipo filogenetico evolutivo, come ad esempio studi riguardanti i cambiamenti della dimensione dei genomi degli organismi presenti all'interno delle divisioni Ascomycota e Basidiomycota, rivelando che l'espansione della taglia del genoma varia in base allo stile di vita dell'individuo, supponendo che i parassiti obbligati hanno genomi piccoli probabilmente per la perdita di geni legati a proteine d'interazione con l'ambiente e una

forte specializzazione verso la cellula ospite, rispetto a quelli dei fitopatogeni come *Fusarium* (Ma *et al.*, 2010) che presenta un genoma molto grande dovuto alla proliferazione delle porzioni non codificanti come introni e elementi mobili (Leducq., 2014).

La comparazione genomica può essere utilizzata anche per analisi di sintenia genica, in quanto nei genomi eucarioti vi sono delle porzioni multi-geniche conservate che tendono a rimanere nelle stesse posizioni del genoma, confrontando invece le parti variabili e come esse cambiano fra specie ed organismi diversi possiamo comprenderne il tipo di evoluzione e il ruolo di queste porzioni nell'adattamento ecologico.

Gli studi di maggiore successo che hanno coniugato le potenzialità della genomica comparata con analisi di tipo funzionale sono stati realizzati su patogeni vegetali come *Fusarium* (Ma *et al.*, 2010), *Magnaporthe* (Xue *et al.*, 2012) e *Colletotrichum* (O'Connell *et al.*, 2012) portando notevoli progressi nell'ambito della patologia vegetale dando ulteriori chiavi di lettura dei meccanismi di patogenicità e di interazione ospite-patogeno di supporto agli studi di tipo sperimentale. Altrettanto importanti sono state le analisi svolte sui generi fungini definiti "benefici" come *Trichoderma* (Kubicek *et al.*, 2011) e *Clonostachys* (Karlsson *et al.*, 2015) portando alla conoscenza dei fattori che concorrono nel processo di micoparassitismo. I risultati di tali studi sono volti ad evidenziare le famiglie geniche espanse o perse coinvolte nei meccanismi di adattamento alle varie nicchie ecologiche sottolineando una elevata variabilità di specializzazione trofica presente sia a livello di specie sia a livello di individuo.

1.3.2 Espansioni geniche in relazione alla specializzazione ecologica

I funghi come gli oomyceti hanno evoluto diversi tipi di meccanismi trofici legati all'habitat in cui vivono e al loro stile di vita come mutualisti, biotrofi, emibiotrofi, necrotrofi e non patogeni come i saprotrofi. Numerosi studi hanno dimostrato che uno dei principali meccanismi di adattamento ecologico fungino è la produzione di piccole proteine extracellulari e metaboliti secondari secreti dall'organismo nell'ambiente circostante (Brown *et al.*, 2012). All'interno del pool proteico secreto vi sono famiglie proteiche che possono agire come molecole segnale che modulano l'interazione fungo-ospite dette effettori, inducendo una risposta nell'ospite vegetale che può andare da un'azione di repressione delle difese vegetali per l'attacco da parte di patogeni ma anche per colonizzarvi le radici come nel caso dei funghi endofiti benefici come *Trichoderma* che

producono le proteine *LysM* (Kubicek *et al.*, 2011), o colonizzarvi le cellule nel caso dei biotrofi e degli emibiotrofi. Ad esempio in tutti gli organismi fitopatogeni sono presenti geni codificanti per Candidate Secreted Effectors Proteins (CSEP) che hanno la funzione di elicitori e che vengono espressi nelle diverse fasi di interazione dell'organismo (biotrofo/necrotrofo). Questi geni sembrano maggiormente espressi in fase di pre-colonizzazione (ad esempio nei patogeni biotrofi ed emibiotrofi quando l'austorio fungino aderisce alla cellula ospite vegetale), secernendo queste proteine i parassiti sono in grado di indurre la repressione delle difese vegetali facilitando la colonizzazione (O'Connell *et al.*, 2012). Oltre alla funzione sopracitata, vi sono anche CSEPs in grado di indurre la risposta di ipersensibilità causando la morte programmata delle cellule dell'ospite da parte dei patogeni necrotrofi come in *Fusarium* (Brown *et al.*, 2010).

In tutti le fasi del ciclo vitale sono prodotti una vasta gamma di enzimi litici coinvolti nella degradazione dei polisaccaridi della parete cellulare vegetale e fungina che, nei patogeni rappresentano la prima barriera che l'organismo si trova a dover affrontare in ogni tipo di interazione, ma anche il meccanismo attraverso il quale il fungo riesce a degradare i composti presenti per poi assimilarli ed utilizzarli. Vengono prodotte anche varie classi di metaboliti secondari come peptaboliti con funzione antibiotica come le gliotossine e le viridine prodotte dalle specie di *Trichoderma* (Mukherjee *et al.*, 2013) ma anche altri che hanno funzione tossica in difesa della pianta ospite come nel caso delle classi di alcaloidi prodotti dalle specie di *Clavicipitaceae* che producono alcaloidi specie-specifici con funzioni diverse a seconda se sono stati prodotti dalle specie biotrofe-parassite (*C. purpurea*) o da quelle simbiotici-obbligati (*E. festucae*) (Schardl *et al.*, 2013).

1.3.3 Famiglie geniche espanse nei micoparassiti e agenti di biocontrollo

All'interno degli organismi fungini agenti di biocontrollo (BCAs) vi sono specie micoparassite appartenenti ai generi *Trichoderma* e *Clonostachys* la cui abilità sembra essere associata a meccanismi di antagonismo, parassitismo, uccisione della preda e nutrimento della biomassa fungina della specie predata (Karlsson *et al.*, 2015). . Questo meccanismo riveste ruoli importanti nell'impiego di questi organismi in agricoltura nella protezione delle specie vegetali combattendo numerosi funghi patogeni, in particolare quelli tellurici come *R. solani* e riuscendo a degradare e colonizzare le durevoli strutture di resistenza melanizzate come gli sclerozi di *Sclerotinia sclerotiorum*, *Macrophomina*

phaseolina e *Verticillium dahliae* (Druzhinina *et al.*, 2011). La vera funzione degli agenti di biocontrollo viene esplicitata dalla colonizzazione delle radici agendo come un simbionte avirulento delle piante, promuovendone la crescita e l'induzione delle risposte di difesa in modo da proteggerle dagli attacchi di patogeni (Druzhinina *et al.*, 2011; Kubicek *et al.*, 2011; Hermosa *et al.*, 2012; Mukherjee *et al.*, 2013).

Durante queste interazioni vengono secreti enzimi legati alla degradazione delle pareti vegetali e fungine (CWDEs) come chitinasi, B-1,3-glucanasi, B-1,6-glucanasi e proteasi, ma anche antibiotici come peptaboliti tipo gliotossine, viridine e 6-pentil-2H-piran-2-one (Lorito *et al.*, 2010; Druzhinina *et al.*, 2011; Mukherjee *et al.*, 2013; Karlsson *et al.*, 2015). Recentemente sono stati sequenziati e annotati diversi genomi delle specie di *Trichoderma* come come *T. virens*, *T. atroviride* e *T. reesei* (Kubicek *et al.*, 2011), *T. hamatum* (Studholme *et al.* 2013) e *T. longibrachiatum* (Xie *et al.*, 2014). Nonostante l'interesse scientifico e economico legato alla comprensione dei meccanismi di azione dei BCAs e la disponibilità delle tecniche a basso costo, solo 3 genomi del genere *Trichoderma* uniti a quello di *C. rosea* (Karlsson *et al.*, 2015) sono state rese pubbliche.

Nel lavoro di Kubicek (2011) sono state comparate tre specie di *Trichoderma*, due micoparassite *T. virens* e *T. atroviride* con la specie saprotrofa *T. reesei* ipotizzando che i caratteri del micoparassitismo in questo genere potessero essere di natura ancestrale e la specie saprotrofa *T. reesei* durante la sua evoluzione avrebbe perso questi caratteri. *T. reesei* infatti mostra una dimensione del genoma inferiore rispetto alle altre due specie *T. virens* e *T. atroviride*, tuttavia presenta espansioni geniche associate al metabolismo dell'azoto. Kubicek e colleghi hanno ipotizzato che tale capacità potesse essere stata acquisita per trasferimento orizzontale (Horizontal gene transfer – HGT) di una nitrato riduttasi proveniente da una specie saprotrofa basidiomicete. Tale fenomeno è stato supposto essere una conseguenza dell'azione micoparassitaria della specie di *Trichoderma* ancestrale nei confronti del basidiomicete. Le famiglie geniche maggiormente espanse nelle specie micoparassite *T. virens* e *T. atroviride* sono quelle coinvolte nella degradazione dei carboidrati e delle proteine presenti nella parete cellulare vegetale e fungina (CWDEs) e quelle legate alla biosintesi dei metaboliti secondari come le polichetide sintasi (PK) entrambe vengono confermate dai risultati ottenuti dall'analisi trascrittomica (Atanasova *et al.*, 2013). Mentre nell'analisi svolta da Karlsson *et al.* (2015) sulle espansioni geniche legate al processo di micoparassitismo di *Clonostachys rosea* sono state riportate le polichetide sintasi come in *Trichoderma* ma in numero più elevato ed in quantità inferiore le chitinasi. Una delle espansioni più significative in *C. rosea* è quella associata alla

produzione di un numero elevato di trasportatori-ABC (ATP-binding cassette) probabilmente coinvolti nella resistenza alla micotossina ZEN (Zearalenone) e a sostanze xenobiotiche. Poichè, *C. rosea* è noto come agente del biocontrollo del Fusarium Head Blight (FHB) per questo motivo durante la sua evoluzione potrebbe aver espanso questa famiglia genica per acquisire resistenza alle specie di *Fusarium* produttrici della micotossina.

1.3.3.1 Cazy (Carbohydrate Active Enzymes)

Il micoparassitismo è un insieme di eventi, tra i quali è presente anche la lisi della parete cellulare fungina. La necessità di degradare i carboidrati presenti nella parete fa sì che si abbia la produzione di cell-wall-degrading enzymes (CWDEs) come gli enzimi chitinolitici, le idrolasi glicosidiche e le β 1-3 glucanasi (Lorito *et al.*, 2010; Kubicek *et al.*, 2011; Daguerre Y. *et al.*, 2014). Questi CWDEs appartengono al gruppo enzimatico dei CaZy (Carbohydrate Active Enzymes) rappresentati da 5 classi ben distinte: le glicoside idrolasi (GH) che agiscono idrolizzando i legami glicosidici; le glicoside transferasi (GT) che catalizzano il trasferimento di molecole di zucchero; le polisaccaride liasi (PL) implicate nella degradazione dei polisaccaridi; le carboidrati esterasi (CE) che promuovono le reazioni di acetilazione degli zuccheri ed infine gli enzimi ad attività ausiliaria (AA) al cui interno sono presenti anche enzimi con attività ligninolitica.

La produzione di CaZy è attiva in tutti gli organismi, anche se maggiore in quelli aventi interazioni di tipo trofico con le cellule vegetali e/o fungine. Quindi si ritrovano anche nei patogeni necrotrofi come *Fusarium* ma anche nei biotrofi come *B. cinerea* e negli emibiotrofi come *Colletotrichum* che produce un elevato numero di pectine liasi che sembrano essere associate ad aumento di virulenza del patogeno nei confronti della pianta ospite. Nel micoparassitismo le classi di CaZy maggiormente espanse sono le glicoside idrolasi (GH) legate alla degradazione della chitina come le endo-B-N-acetilglucosaminidasi (GH18), le B-1,3-glucanasi, ma anche classi di chitinasi e chitosanasi sono presenti nelle specie di *Trichoderma* (Kubicek *et al.*, 2011). Le GH18 sembrano avere ruoli chiave anche in altri agenti di biocontrollo come *C. rosea*. Le GH18 sono costituite da tre sottogruppi: due aventi funzioni di eso e endochinasi ed uno di “killer-toxin chitinasi like” supponendo che l’azione di quest’ultimo sia quella di rendere permeabile la parete cellulare della preda fungina facilitando la penetrazione delle tossine prodotte dal micoparassita (Karlsson *et al.*, 2015). La presenza di geni codificanti le

GH18 con funzione di 'azione di "killer-toxin chitinasi like" sono stati trovati anche in *T. virens* e in *T. atroviride* ipotizzando che la loro presenza sia dovuta alla dipendenza della produzione di metaboliti tossici durante le fasi di micoparassitismo, anche se è stato più volte marcato che il ruolo principale delle GH18 negli isolati di *Trichoderma* sia quello di endochitinasi (Kubicek *et al.*, 2011; Seild- Seiboth *et al.*, 2014).

1.3.3.2 Peptidasi

Le peptidasi, chiamate anche proteasi, sono enzimi proteolitici che assieme alle glucanasi e chitinasi (CaZY) contribuiscono alla degradazione della parete cellulare, agendo appunto sulle proteine di parete dell'ospite (Suarez *et al.*, 2007; Kubicek *et al.*, 2011; Antanasova *et al.*, 2013). Negli isolati di *Trichoderma* agenti di biocontrollo sono secrete durante la fase di antagonismo e sembrano avere un ruolo anche nel degradare e disattivare enzimi prodotti dai patogeni durante il processo di infezione delle specie vegetali (Suarez *et al.*, 2007).

Inizialmente le peptidasi erano suddivise in cinque famiglie in base all'aminoacido catalitico: aspartato, cisteina, treonina, serina e metallo (queste ultima utilizzano un metallo come cofattore); con l'aumento della disponibilità di genomi sequenziati sono state scoperte nuove famiglie che hanno portato ad una riclassificazione di quelle già conosciute ma soprattutto ne sono state aggiunte delle nuove come le peptidasi glutammiche e le asparagine.

Nei genomi di *Trichoderma* sequenziati e disponibili, sono state osservate espansioni geniche a carico delle aspartato e serine peptidasi. Il ruolo di tali proteine è stato confermato da analisi di trascrittomiche delle specie micoparassite nei confronti di *R. solani* (Antanasova *et al.*, 2013).

In vari isolati micoparassiti di *Trichoderma* sono state identificate le famiglie di proteasi come le aspartato peptidasi associate a geni isolati in isolati di *T. harzianum* e in *T. asperellum* sono stati isolati geni associati ad un'elevata produzione di aspartato peptidasi volte ad aumentare la colonizzazione delle radici vegetali, mentre in un altro ceppo di *T. harzianum* la funzione delle serine peptidasi presenti sembrava essere quella nematocida (Suarez *et al.*, 2007).

1.3.3.3 Classi proteiche coinvolte nella sintesi di metaboliti secondari

Gli organismi producono vari gruppi di metaboliti secondari come i polichetidi prodotti dai geni codificanti per la polichetide sintasi (PKS), peptidi prodotti dalle peptide sintasi non ribosomali (NRPS), alcaloidi da quelli delle dimetilallil triptofano sintasi (DMATS) e i metaboliti volatili come i terpeni delle terpene sintasi (TS). Gli organismi patogeni utilizzano i metaboliti secondari anche per reprimere le difese della pianta e poterla colonizzare e nutrirsi, altri invece producono tossine e antibiotici come meccanismi di difesa e attacco. Sia le specie micoparassite di *Trichoderma* prese in considerazione nell'analisi comparata di Kubicek *et al.*, 2011, sia in *C. rosea* sono mostrate espansioni delle famiglie proteiche associate alle polichetide sintasi (PKS) e peptide sintasi non ribosomali (NRPS), tuttavia la loro funzione specifica non del tutto chiara anche se l'ipotesi maggiormente avvalorata è quella di avere un ruolo nella biosintesi di composti antifungini. In *T. virens* sono presenti un alto numero di NRPS, ma soprattutto le famiglie proteiche associate al citocromo P450 ossidasi che prende parte alla biosintesi di numerosi metaboliti secondari.

2. Scopo del lavoro

Il progetto di sequenziamento dei genomi di *T. harzianum* (T6776) e *T. gamsii* (T6085) nasce dall'esigenza di approfondire le conoscenze genetiche alla base dello stile di vita (micoparassitismo e interazione con la pianta) degli organismi appartenenti a questo raggruppamento tassonomico.

Tale studio ha l'obiettivo di fornire una piattaforma utile per futuri studi di genomica funzionale quali analisi di trascrittomica, metabolomica e proteomica.

Il seguente lavoro si è posto due obiettivi principali:

- sequenziare, assemblare e annotare per la prima volta i genomi di isolati rappresentativi appartenenti alle specie *Trichoderma harzianum* e *Trichoderma gamsii*.
- effettuare un'analisi di comparazione genomica per identificare, in modo preliminare, le famiglie proteiche coinvolte nello stile di vita dei due isolati, focalizzandosi in misura maggiore sulle classi enzimatiche coinvolte nei processi di micoparassitismo e nell'azione di biocontrollo relative ai CaZy (*Carbohydrate-Active enZYmes*).

3. Materiali e metodi

3.1 *Isolati fungini*

Gli isolati fungini, oggetto di studio della presente tesi, sono *T. harzianum* T6776 e *T. gamsii* T6085 appartenenti alla collezione fungina depositata presso il Laboratorio di Micologia Fitopatologica, Dipartimento di Scienze Agrarie, Alimentari e Agro-ambientali dell'Università di Pisa.

L'isolato T6776 di *T. harzianum*, isolato per la prima volta in Italia dal terreno di un pescheto a Colignola (Pisa) mentre l'isolato *T. gamsii* T6085 proviene da un terreno non coltivato in Crimea, Ucraina. La storia dei due isolati è dettagliatamente riportata nell'introduzione.

3.2 *Estrazione DNA genomico*

Il DNA è stato estratto da micelio fungino fatto crescere in piastre Petri contenenti PDA (*Potato Dextrose Agar*) sopra un dischetto di cellophane sterile per 3/4 giorni a 25°C in assenza di luce.

L'estrazione è stata eseguita secondo il protocollo del kit DNeasy Plant Mini (Qiagen Inc., Valencia, CA, USA); la qualità dell'estrazione è stata osservata tramite corsa elettroforetica su gel d'agarosio (1%) in TBE (0,5X) e quantificata con misurazione spettrofotometrica rilevando l'assorbanza a 260/280 nm.

L'identità degli isolati prescelti è stata confermata amplificando e sequenziando la regione ribosomale ITS (*Internal Transcribed Spacer*). Per amplificare la regione di rDNA contenente l'ITS1 e ITS2 sono stati utilizzati i primers ITS5 (5'-GGAAGTAAAAGTCGTAACAAGG-3') e ITS4 (5'-TCCTCCGCTTATTGATATGC-3') (White *et al.*, 1990) (Fig.3.1).

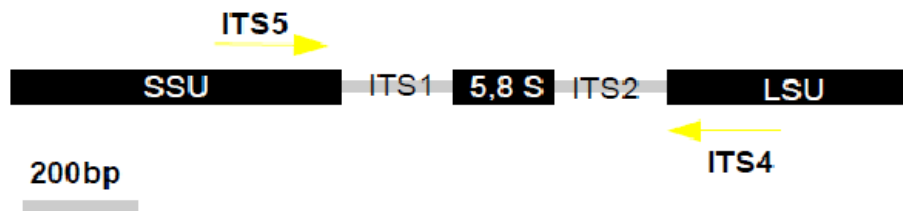


Figura 3.1. Localizzazione dei primers ITS5 e ITS4 impiegati per l'amplificazione delle regioni ribosomiali ITS1, 5,8S e ITS2.

L'amplificazione è stata condotta in condizioni standard mediante GeneAmp PCR System 2400 (Perkin Elmer) con una denaturazione iniziale di 1 min a 94°C, seguita da 30 cicli di amplificazione che prevedono:

- denaturazione 30 sec a 94°C
- primer annealing 1 min a 54°C
- estensione 1 min a 72°C

dopo un'estensione finale di 4 min a 72°C, l'amplificato viene conservato a 4°C.

Prima di essere inviati a sequenziare, i prodotti di PCR sono stati purificati tramite il kit QIAquick (Qiagen Inc., Valencia, CA, USA) e visualizzati su gel di agarosio (1%) in TBE (0,5X); come marcatore di qualità è stato utilizzato il DNA LADDER 1Kb della Promega.

Il sequenziamento dei prodotti di PCR è stato condotto mediante un sequenziatore Applied Biosystem Prism® 3100 DNA sequencer (Perkin Elmer/ABI) attraverso il BigDye™ terminator v3.1 Cycle Sequencing Ready Reaction Kit. I campioni sequenziati sono stati processati presso la Bio Molecular Research (BMR), Servizio Sequenziamento – CRIBI, dell'Università di Padova.

L'identità delle sequenze è stata determinata inserendo i prodotti sequenziati all'interno del database TrichOKEY v.2.0 presente nel sito della "Sottocommissione Internazionale di *Trichoderma* e *Hypocrea*" (ISTH, www.isth.info) (Druzhinina and Kopchinski, 2006) oppure tramite GenBank (<http://www.ncbi.nlm.nih.gov>).

Per il sequenziamento genomico sono stati utilizzati 50 µl di DNA con concentrazione di 20ng/µl.

3.3 Sequenziamento

Il sequenziamento è stato effettuato presso il Génome Québec Innovation Centre (McGill University), Québec, Canada. Sono state sequenziate librerie genomiche con insert-size di 700 bp (Fig.3.2), costruite utilizzando TruSeq™ RNA e DNA Sample Preparation Kits (Illumina®) e sequenziate con la tecnologia Miseq PE250 Illumina®. Il processo prevede la frammentazione del DNA estratto tramite sonificazione, i frammenti ottenuti vengono separati tramite gel e la banda relativa alla lunghezza desiderata (insert-size) viene eluita. I frammenti selezionati vengono ligati ad adattori di sequenza nota che verranno utilizzati per una serie di amplificazioni e per il sequenziamento di ogni singolo frammento.

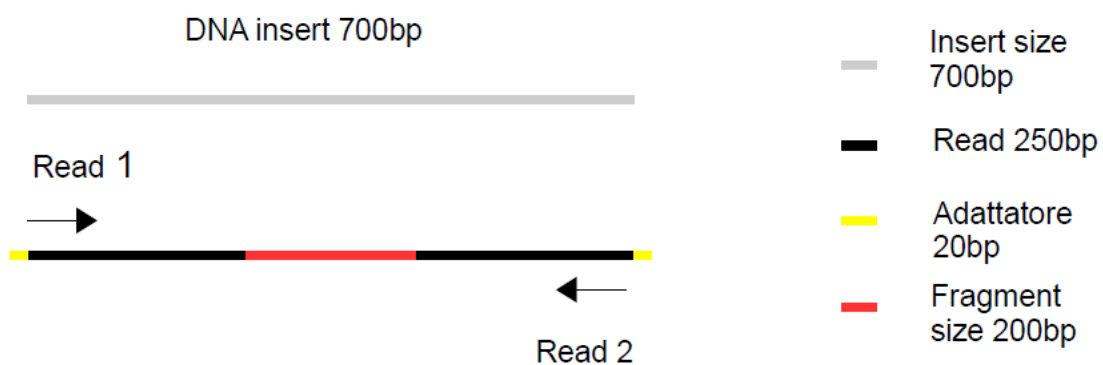


Figura 3.2. Sono rappresentate le sequenze Mate pair ends Illumina di 250bp; con Insert size di 700bp; fragment-size di 200bp e alle estremità vi sono gli adattatori di 20bp.

La tecnologia scelta ha il vantaggio di produrre sequenze più lunghe, Mate pair ends (250bp x2), rispetto ad altre piattaforme Illumina (30-100 nucleotidi), mantenendo costi relativamente bassi e producendo un'elevata quantità di dati.

3.3.1 Valutazione del sequenziamento

Dal servizio di sequenziamento sono stati ottenuti, per ogni genoma, due file, uno per le sequenze forward e uno per quelle reverse. I file, in formato *fastQ*, contengono per ogni nucleotide un valore che indica la qualità del segnale della singola base.

Per valutare se il sequenziamento è andato a buon fine, vengono analizzate in modo quantitativo e qualitativo, le sequenze prodotte:

- nella prima viene calcolata la copertura media del genoma o “genome coverage” (W) mediante la seguente formula:

$$W=NL/G$$

dove N rappresenta il numero di sequenze ottenute dal sequenziamento, L la loro lunghezza espressa in coppie di basi (bp), entrambe rapportate a G che corrisponde alle dimensioni medie dei genomi presenti in letteratura di organismi dello stesso genere o filogeneticamente vicini a *Trichoderma*.

La copertura media del genoma esprime quante volte il numero totale di basi del genoma è rappresentato dai dati ottenuti attraverso il sequenziamento (es. 50X di copertura indica che le informazioni ottenute sono 50 volte le dimensioni del genoma oggetto di studio) (Schatz *et al.*, 2012). La copertura ottimale varia in base alla dimensione e alla natura del genoma dell'organismo che viene sequenziato.

- la seconda analisi viene eseguita attraverso il software FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Si tratta di un'interfaccia grafica che, su base statistica, permette di osservare la qualità delle sequenze valutando il punteggio della qualità (espresso in Phred) di ogni base presente nella sequenza e il punteggio complessivo di tutte le sequenze; l'andamento della distribuzione delle basi all'interno la sequenza; la percentuale di contenuto in guanina e citosina (GC); la lunghezza media delle sequenze; il numero di basi non sequenziate (N) contenute nelle sequenze; la presenza/assenza degli adattatori e la quantità di sequenze duplicate.

3.4 Assemblaggio del genoma

I genomi sono stati assemblati utilizzando il programma Velvet 1.2.08, algoritmo sviluppato sull'approccio dei Brujin Graphs impiegato per assemblaggi di genomi *de novo* con short reads (Zebrino and Birney, 2008). Questo metodo prevede la suddivisione delle sequenze in “short words”, oligonucleotidi chiamati k-mers, i quali si andranno ad appaiare ad altri k-mers per omologia di prefisso/suffisso andando a formare i nodi all'interno di un grafo. Seguendo il percorso segnato dai nodi presenti nel grafo si ottengono lunghi frammenti di sequenze consenso detti contigs oppure scaffolds, nel caso in cui le informazioni relative alle sequenze appartenenti allo stesso frammento vengano utilizzate

per connettere i contigs fra di loro formando sequenze ancora più lunghe.

Velvet è composto da due parti denominate VelvetH e VelvetG, dove VelvetH viene impiegato come primo step per suddividere il genoma in K-mers di lunghezza nota (nel caso specifico tra 29 e 69). Il secondo step viene eseguito da VelvetG dove vengono impostati i parametri per l'assemblaggio. *In primis* vengono forniti come input i dati provenienti dall'analisi di VelvetH, dopodiché verranno settati l'insert-size, la copertura media attesa del genoma, la soglia di copertura minima del genoma in modo da eliminare zone che hanno copertura minore del valore impostato dall'utente e la lunghezza minima degli scaffolds.

3.4.1 Valutazione della completezza dell'assemblaggio

La completezza del genoma assemblato è stata valutata tramite il programma CEGMA versione 2.4 (Parra *et al.* 2007). Si tratta di un metodo computazionale, basato sul Hidden Markov Models. L'Hidden Markov Models o Modello di Markov è un modello statistico che viene utilizzato per la ricerca e predizione di geni che si basa sull'identificazione di uno “stato”, che potrebbe essere un esone, introne o sito di splicing, basandosi su informazioni biologiche e sul miglior punteggio di probabilità ottenuto.

CEGMA utilizza questo modello per individuare la struttura genica all'interno della sequenza genomica, fornendo una stima, espressa in percentuale, della completezza dell'assemblaggio sulla base della presenza nel genoma e dell'eventuale frammentazione di 458 geni ortologhi conservati negli organismi eucariotici.

3.5 Predizione e annotazione genica

La predizione della struttura genica è stata eseguita in maniera automatica attraverso MAKER2 (Holt and Yandell, 2011). Maker è una pipe-line specifica per l'annotazione di genomi eucariotici e si avvale di diversi software che operano su modelli e informazioni differenti.

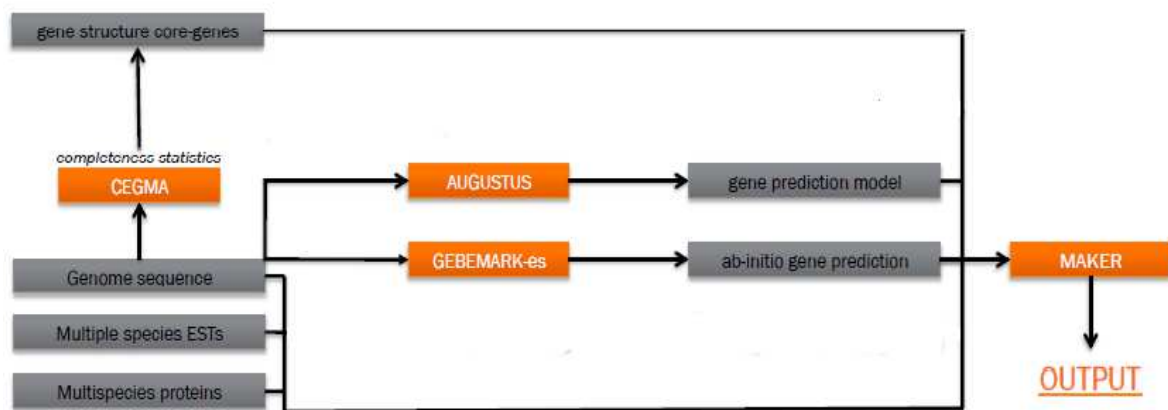


Figura 3.3. Maker2 pipe-line. Sono rappresentati i passaggi e i programmi utilizzati per la predizione del contenuto genico. In arancione sono evidenziati i programmi, mentre in grigio sono evidenziate le informazioni utilizzate dai programmi e i loro output.

Per la predizione genica di entrambi i genomi di *Trichoderma* è stata seguita la pipe-line rappresentata nella Figura 3, utilizzando i programmi GeneMark-ES e AUGUSTUS. GeneMark-ES v.2 (Ter-Hovhannisyanyan *et al.* 2008) è un algoritmo che opera in maniera indipendente, identifica i geni codificanti in modo probabilistico secondo il modello di Markov basandosi solo sulla sequenza genomica. AUGUSTUS (Stanke *et al.* 2004) invece, segue un approccio più robusto che combina la ricerca probabilistica con informazioni biologiche ottenute sperimentalmente o provenienti da organismi evolutivamente vicini a quello oggetto di studio. Si tratta di un programma che richiede un modello per identificare i geni putativi presenti nel genoma. Nel caso di *T. harzianum* e *T. gamsii*, non essendo disponibili evidenze biologiche come sequenze di trascritti o proteine derivanti da analisi omiche, sono stati utilizzati i modelli predittivi di *Fusarium graminearum* e di *Magnaporthe grisea*.

Inoltre, Maker permette di aggiungere ulteriori informazioni affinché si possa ottenere una predizione genica molto accurata.

Al fine di migliorare la predizione genica sono state utilizzate evidenze quali EST (expressed sequence tags) di *T. harzianum*, *T. virens*, *T. atroviride* disponibili dal sito del DOE Joint Genome Institute (JGI) (<http://genome.jgi.doe.gov/programs/fungi/index.jsf>) e sequenze di proteine fungine disponibili su SWISS-PROT, database di proteine curate manualmente (Bairoch *et al.* 2000).

3.6 Comparazione genomica

I due genomi oggetto di studio sono stati comparati con quelli di altri organismi fungini. La scelta degli organismi da comparare è stata effettuata considerando sia le caratteristiche biologiche che le distanze evolutive, scegliendo organismi con comportamenti biologici differenti ma evolutivamente vicini a *Trichoderma*, in modo da riuscire a mettere in evidenza le famiglie geniche espanse o ristrette dei due isolati durante il loro adattamento.

3.6.1 Scelta degli organismi

Inizialmente sono state inserite le altre specie di *Trichoderma* sequenziate e disponibili (*T. virens*, *T. atroviride*, *T. reesei*) dopodiché come out group sono stati selezionati altri Sordariomiceti tra i quali agenti causali di malattie come *Fusarium graminearum*, *Fusarium oxysporum*, *Fusarium verticillioides*, *Magnaporthe oryzae*, *Claviceps purpurea* e *Colletotrichum fioriniae* ma anche organismi modello per lo studio dell'ecologia fungina come *Neurospora crassa*.

Gli aspetti di primaria importanza che hanno guidato la scelta degli organismi da inserire nella comparazione genomica sono stati il comportamento ecologico e lo stile di vita.

Sono stati inseriti organismi saprotrofi come *Neurospora crassa* e *Trichoderma reesei*, emibiotrofi come *Magnaporthe oryzae* e *Colletotrichum fioriniae*, un biotrofo obbligato come *Claviceps purpurea*, le tre specie necrotrofe appartenenti al genere *Fusarium* ed infine gli organismi con comportamenti simili agli isolati oggetto di studio come *T. atroviride* e *T. virens*, noti agenti di biocontrollo, e *Clonostachys rosea* micoparassita, endofita e agente di biocontrollo.

3.6.2 Analisi comparata

L'analisi di comparazione genomica è stata condotta a partire dall'insieme delle proteine putative predette (chiamato anche proteoma) degli organismi, le quali sono state filtrate in modo da ottenere solo la porzione di proteine secrete. Il secretoma è l'insieme dei geni che codificano per le proteine che vengono secrete nell'ambiente extracellulare da parte di un organismo (Lorito *et al.*, 2010).

A partire dalle famiglie proteiche secrete, da parte di tutti gli organismi, è stato possibile mettere in evidenza quelle maggiormente prodotte dal genere *Trichoderma* ponendo

un'attenzione particolare su quelle coinvolte nel processo di micoparassitismo gli isolati oggetto della presente tesi.

Genus	Specie	Strain	Code	Accession	Source	Reference
Neurospora	crassa	or74a	NCRA	AABX00000000.2	Broad	Galagan JE, <i>et al.</i> Nature 2003
Magnaporthe	oryzae	70-15	MORY	AACU00000000.3	Broad	Xue M, <i>et al.</i> PLoS Genet 2012
Colletotrichum	fiorinae	PJ7	CFIO	JARH01000000.1	NCBI	Baroncelli R, <i>et al.</i> Genome Announc, 2014
Claviceps	purpurea	20,1	CPUR	CAGA00000000.1	NCBI	Schardl CL, <i>et al.</i> PLoS Genet 2013
Trichoderma	atroviride	IMI 206040	TATR	ABDG00000000.2	NCBI	Kubicek CP, <i>et al.</i> Genome Biol 2011
Trichoderma	gamsii	T6085	TGAM	JPDN00000000.1*	private	presente lavoro
Trichoderma	reseei	QM6a	TRES	AAIL00000000.2	NCBI	Martinez D, <i>et al.</i> Nat Biotechnol 2008
Trichoderma	harzianum	T6776	THAR	JOKZ00000000.1*	private	presente lavoro
Trichoderma	virens	Gv29-8	TVIR	ABDF00000000.2	NCBI	Kubicek CP, <i>et al.</i> Genome Biol 2011
Fusarium	verticillioides	7600	FVER	AAIM00000000.2	Broad	Ma LJ, <i>et al.</i> Nature 2010
Fusarium	oxysporum	FOL 4287	FOXY	AAXH00000000.1	Broad	Ma LJ, <i>et al.</i> Nature 2010
Fusarium	graminearum	PH1	FGRA	AACM00000000.2	Broad	Ma LJ, <i>et al.</i> Nature 2010
Clonostachys	rosea	BAFC3874	CROS	*	private	Karlsson M, <i>et al.</i> Genome Biol 2015

Tabella 3.1. Organismi impiegati nell'analisi di comparazione genomica. Vi sono indicati il nome dell'organismo, il ceppo, il codice utilizzato nelle analisi, l'accession number per i dati relativi ad esso, l'organo dove sono stati depositati i dati e la pubblicazione di riferimento.

3.7 Analisi del secretoma

L'analisi comparata ha preso inizio con l'identificazione e lo studio delle proteine putative secrete di ogni organismo. Per identificare i secretomi abbiamo utilizzato WoLF PSORT (Horton *et al.* 2007), un programma che, attraverso la composizione della porzione aminoacidica della proteina, è capace di predirne la localizzazione cellulare basandosi sull'allineamento di sequenze. Per la predizione si utilizzano anche altri database come UniProt e Gene Ontology (GO) (Horton *et al.* 2007).

Le sequenze delle proteine secrete dagli organismi oggetti di studio sono state successivamente analizzate con InterProScan (Quevillion E. *et al.* 2005), un programma atto ad identificare i domini proteici conservati presenti nei prodotti genici secreti. Ad ogni motivo proteico conservato corrisponde un profilo IPR il quale identificherà la famiglia proteica corrispondente e i siti funzionali del gene, fornendo valide informazioni sul ruolo biologico svolto dalla proteina stessa.

Specifiche famiglie proteiche, con un noto ruolo nell'interazione con l'ambiente, sono state annotate manualmente sulla base dei domini conservati, avvalendosi del database InterPro (<http://www.ebi.ac.uk/interpro/>) (Haunter S. *et al.* 2009; 2011).

Al fine di analizzare le basi genomiche del comportamento e dell'adattamento evolutivo delle due specie di *Trichoderma*, lo studio delle proteine secrete si è focalizzato in misura maggiore sulla caratterizzazione dei CAZy (*Carbohydrate-Active enZYmes*), una classe enzimatica coinvolta nel processo di micoparassitismo e nell'azione di agente di biocontrollo (BCA), già nota in *Trichoderma* (Kubicek CP *et al.* 2011; 2014; Xie *et al.* 2014).

I CAZy sono stati caratterizzati mediante annotazione automatica con dbCAN (Yin *et al.* 2012), supportata da un'analisi manuale sul database dei CaZy (<http://www.cazy.org/>).

4. Risultati e Discussione

4.1 Valutazione della qualità del DNA genomico

La qualità del DNA genomico degli isolati T6776 e T6085 è stata valutata osservando l'integrità delle bande tramite corsa elettroforetica su gel di agarosio (1%) in TBE (0.5X), mentre la purezza e la concentrazione sono state quantificate attraverso la lettura delle assorbanze allo spettrofotometro. La concentrazione del DNA a doppio filamento è stata determinata con la lettura dell'assorbanza a 260nm, mentre la purezza è stata determinata attraverso i rapporti tra l'assorbanza a 260nm e quella a 280nm, per valutare eventuali contaminazioni da proteine e dal rapporto tra 260nm e 320 nm per contaminazioni da solventi contenenti carboidrati e fenoli. Per entrambi gli isolati sono stati ottenuti valori positivi della concentrazione, dei rapporti A260nm/ A280nm, che rientrano nel range di 1.6-2, e del rapporto A260nm/ A320nm maggiore di 2.2, il valore soglia oltre il quale viene esclusa la possibile contaminazione da fenoli (Tab. 4.1)

Isolato	ng/ μ l	260/280	260/320
<i>T. harzianum</i> T6776	24.3	1.86	2.65
<i>T. gamsii</i> T6085	26.1	1.82	2.95

Tabella 4.1. Qualità del DNA genomico degli isolati *T. harzianum* T6776 e *T. gamsii* T6085. Sono riportati il valore della concentrazione (ng/ μ l), il rapporto dell'assorbanza 260nm/280nm e il rapporto delle assorbanze 260nm/320nm.

L'identità degli isolati è stata confermata amplificando e sequenziando il locus genico della regione ribosomale ITS (*Internal Transcribed Spacer*), contenente ITS1, 5.8S e ITS2.

I prodotti di PCR, prima di essere inviati a sequenziare, sono stati purificati e quantificati su gel d'agarosio (1%) in TBE (0.5X) ottenendo frammenti di peso molecolare di circa 700bp. I risultati del sequenziamento hanno confermato l'identità degli isolati di interesse.

4.2 Valutazione del sequenziamento

Il sequenziamento delle librerie genomiche degli isolati fungini ottenuto mediante MiSeq

PE250 Illumina® ha prodotto circa $13 \cdot 10^6$ sequenze per *T. harzianum* e $15 \cdot 10^6$ per *T. gamsii*. La lunghezza delle sequenze prodotte è di 250 bp.

Considerando le informazioni presenti in letteratura relative ai genomi delle altre specie di *Trichoderma* sequenziate, è possibile stimare che le dimensioni del genoma degli organismi appartenenti a questo gruppo tassonomico sia di circa 40 Mb.

Il valore o il livello di copertura del genoma è un parametro estremamente importante, in quanto fattore limitante per un buon assemblaggio. Il valore di copertura che andiamo a stimare si riferisce ad una media rapportata su tutto il DNA totale sequenziato. Considerando che all'interno del genoma vi sono regioni altamente ripetute (come ad esempio il cluster ribosomiale o il DNA mitocondriale) il valore reale di copertura delle sequenze uniche diventa nettamente inferiore a quello calcolato. Più la copertura è elevata più è possibile arrivare ad ottenere un genoma finito: privo di gaps, mismatch e costituito da unici scaffolds orientati e assemblati a livello di cromosomi. Ad oggi, per i Sordariomycetes di cui è stato sequenziato il genoma, assemblamenti finiti a livello di cromosomi sono stati ottenuti solo in *M. grisea* e in *N. crassa*.

Con i dati ottenuti dal sequenziamento e dalla letteratura, è stato possibile calcolare la copertura media dei genomi (W) oggetto di studio, ottenendo una copertura di 80X per l'isolato *T. harzianum* T6776 e 90X per *T. gamsii* T6085.

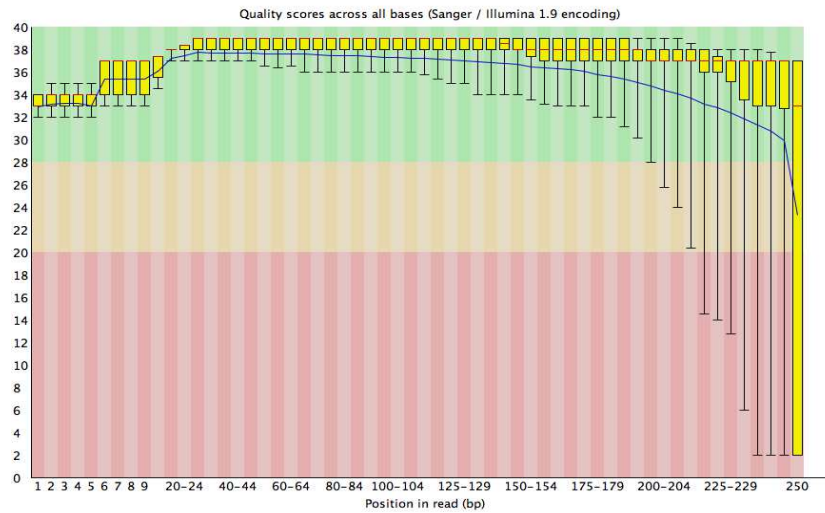
Le coperture ottenute per entrambi gli isolati di *Trichoderma* sono più che soddisfacenti per lo scopo del lavoro prefissato nella presente tesi. Per un'analisi di tipo comparato, al fine di studiare il contenuto genico presente in un organismo, non è necessario arrivare a costituire genomi finiti ma è sufficiente produrre frammenti (contigs, superconting oppure scaffolds) di dimensioni tali a garantire la copertura e l'integrità delle zone codificanti (Chain *et al.*, 2009). Utilizzando la stessa tecnica di sequenziamento, sono stati prodotti genomi di buona qualità anche con coperture inferiori a 50X come nel caso di *Colletotrichum fioriniae* (Baroncelli *et al.*, 2014). Nonostante ciò, la copertura di 50X sembra essere considerata come la soglia minima da raggiungere per ottenere assemblaggi genomici di qualità sufficiente per effettuare analisi sul contenuto genico.

La qualità del sequenziamento è stata determinata attraverso FastQC mediante analisi delle sequenze presenti nei file forward e reverse di entrambi i genomi. I parametri presi in considerazione sono stati: i) la qualità delle basi nelle sequenze ii) il punteggio di qualità dell'insieme di sequenze espresso in phred; iii) la percentuale del contenuto in guanina e citosina (Fig. 4.1-4.4). I valori delle sequenze degli isolati oggetto di studio mostrano una buona qualità delle basi con un punteggio di phred maggiore di 38. Questo valore sta ad

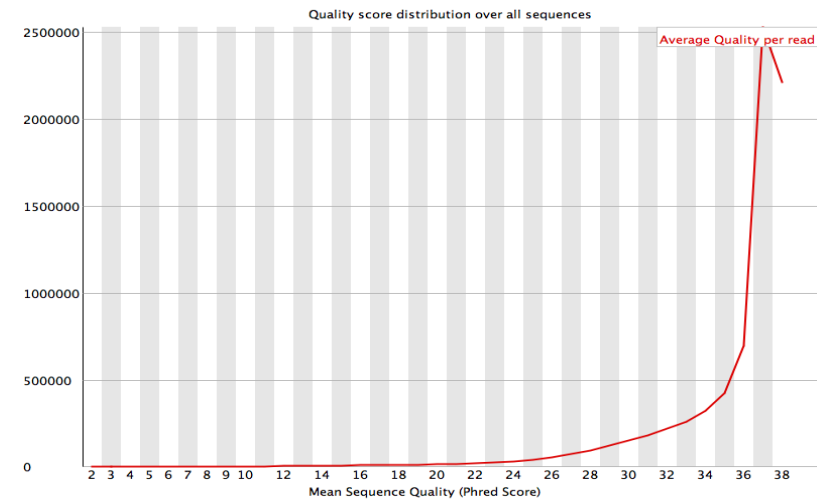
indicare che la probabilità di errore di chiamata delle basi nelle sequenze è basso, quindi minore sarà la presenza di errori dovuti al sequenziamento.

I grafici del contenuto in CG per entrambi i file degli isolati, presentano un unico picco, permettendo di escludere la presenza di contaminanti.

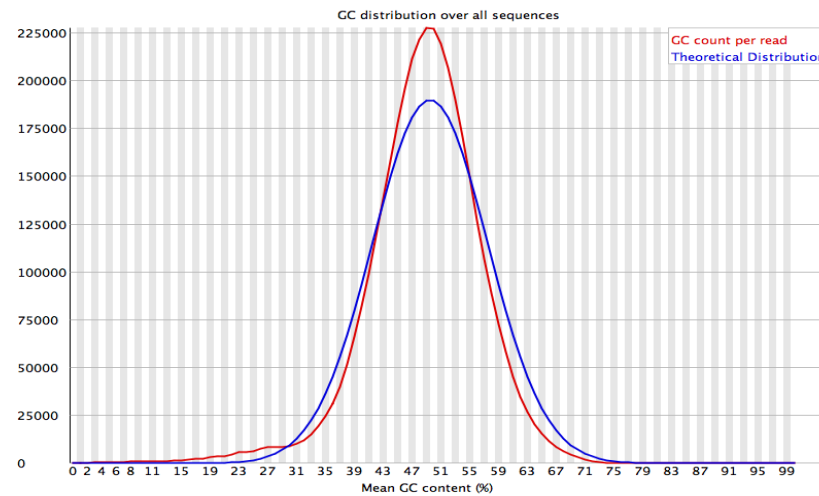
Nel complesso, considerando i valori ottenuti dai dati risultanti dalla tecnologia di sequenziamento scelto, potremmo aspettare un livello di assemblaggio, per entrambi i genomi, tale da coprire le esigenze delle analisi future da condurre.



A

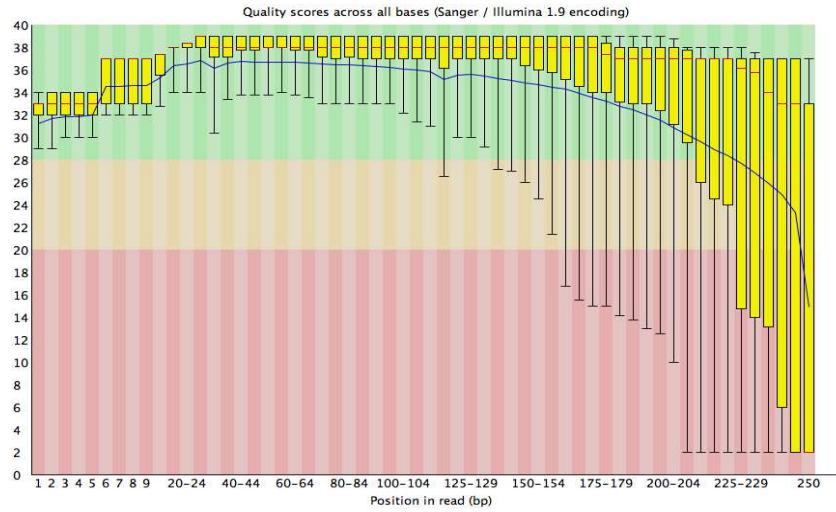


B

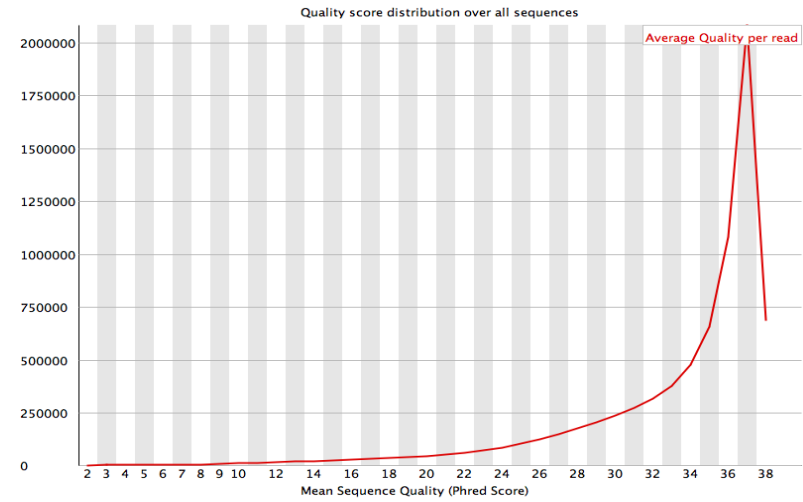


C

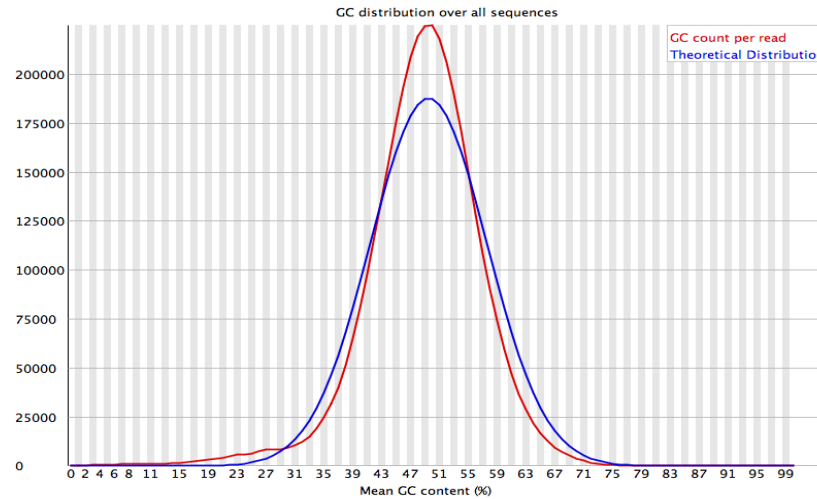
Figura 4.1. Output visivo delle sequenze reverse di T6085 con FastQC. A) Qualità della sequenza per base. La posizione delle basi nelle sequenze sono mostrate sull'asse X e il punteggio di qualità (Q score) vengono visualizzati sull'asse Y; B) Punteggio della qualità delle sequenze. La media è riportata sull'asse X; C) Contenuto in GC delle sequenze. La media del contenuto in GC (%) è riportata sull'asse X. Mostra la distribuzione di GC in tutte le sequenze.



A

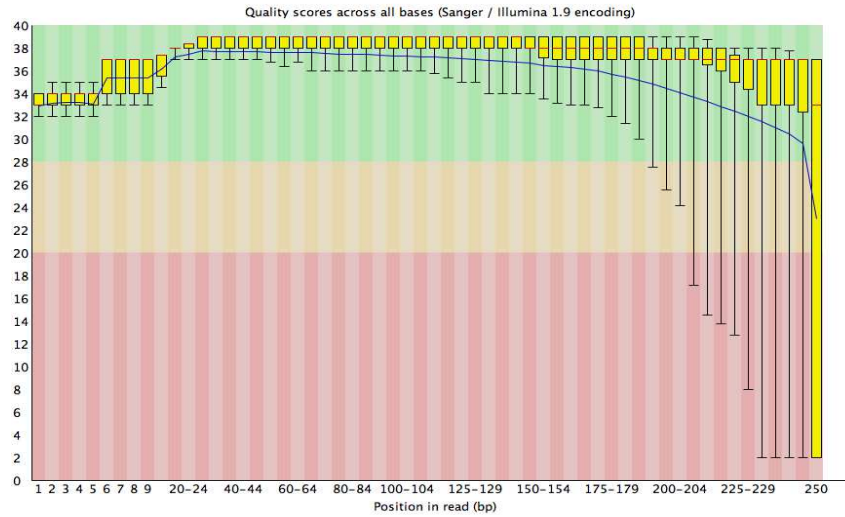


B

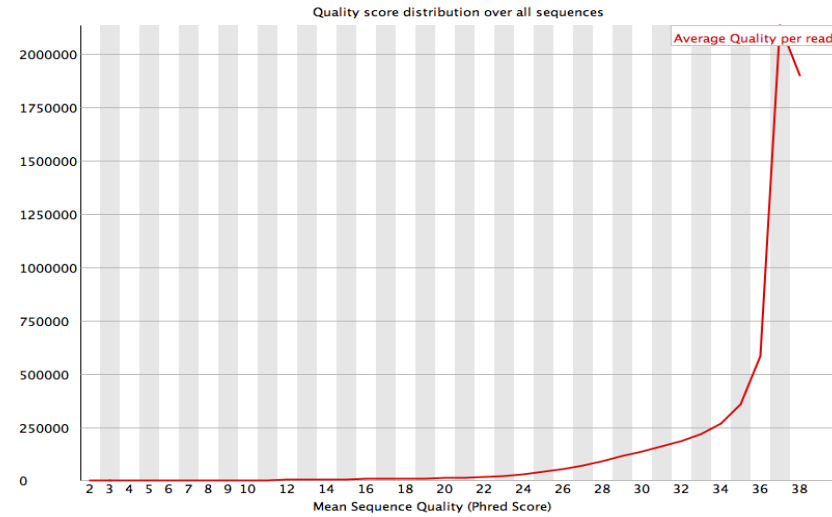


C

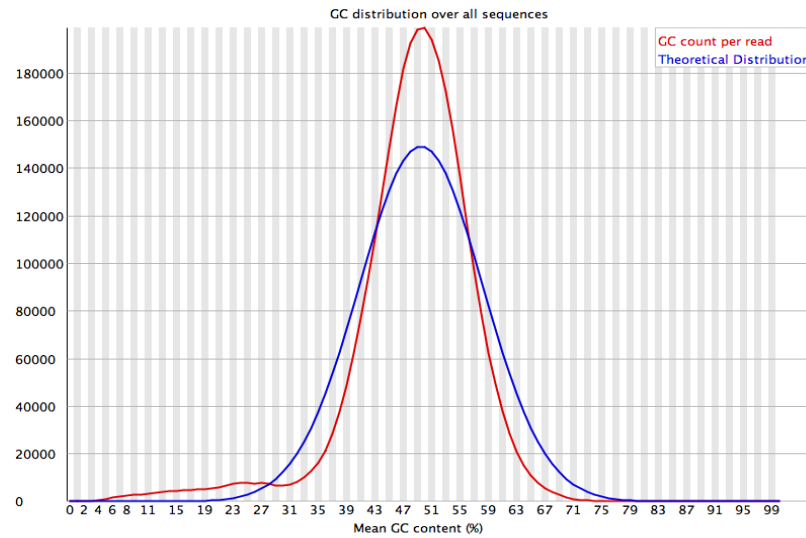
Figura 4.2. Output visivo delle sequenze foward di T6085 con FastQC. A) Qualità della sequenza per base. La posizione delle basi nelle sequenze sono mostrate sull'asse X e il punteggio di qualità (Q score) vengono visualizzati sull'asse Y; B) Punteggio della qualità delle sequenze. La media è riportata sull'asse X; C) Contenuto in GC delle sequenze. La media del contenuto in GC (%) è riportata sull'asse X. Mostra la distribuzione di GC in tutte le sequenze.



A

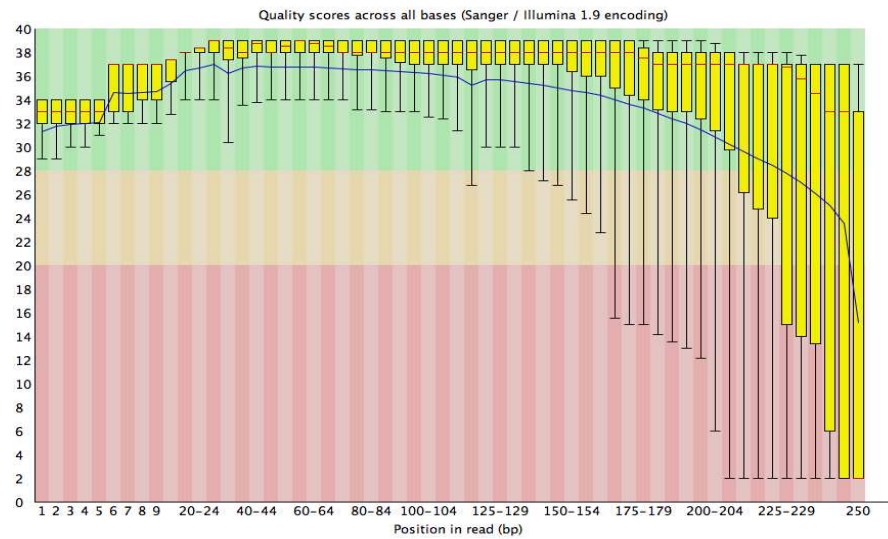


B

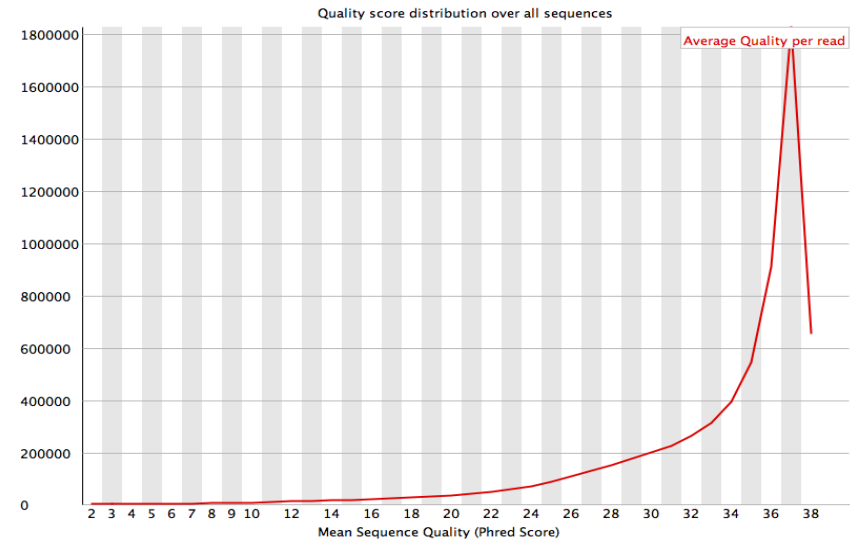


C

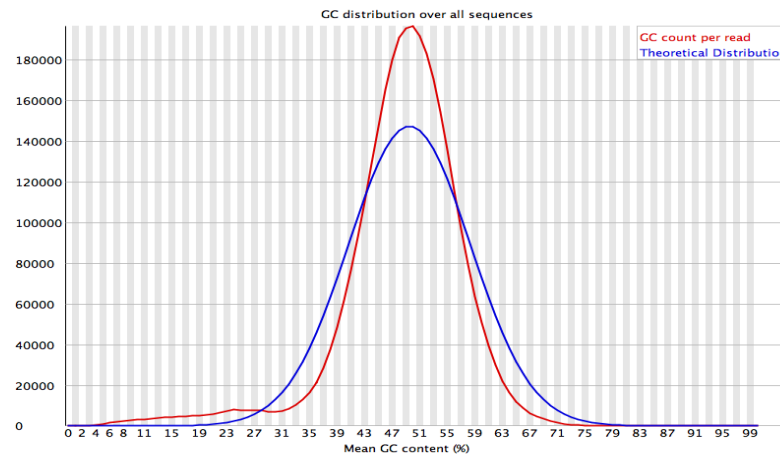
Figura 4.3. Output visivo delle sequenze reverse di T6776 con FastQC. A) Qualità della sequenza per base. La posizione delle basi nelle sequenze sono mostrate sull'asse X e il punteggio di qualità (Q score) vengono visualizzati sull'asse Y; B) Punteggio della qualità delle sequenze. La media è riportata sull'asse X; C) Contenuto in GC delle sequenze. La media del contenuto in GC (%) è riportata sull'asse X. Mostra la distribuzione di GC in tutte le sequenze.



A



B



C

Figura 4.4. Output visivo delle sequenze forward di T6776 con FastQC. A) Qualità della sequenza per base. La posizione delle basi nelle sequenze sono mostrate sull'asse X e il punteggio di qualità (Q score) vengono visualizzati sull'asse Y; B) Punteggio della qualità delle sequenze. La media è riportata sull'asse X; C) Contenuto in GC delle sequenze. La media del contenuto in GC (%) è riportata sull'asse X. Mostra la distribuzione di GC in tutte le sequenze.

4.3 Valutazione dell'assemblaggio

L'obiettivo principale dell'assemblaggio di un genoma è quello di riuscire ad unire più sequenze possibili tra loro andando a formare delle sequenze più lunghe, contig o scaffolds, che riescano a coprire l'intera del genoma ed evitando così errori e gap fisici. Le sequenze genomiche ottenute sono state assemblate con il software Velvet 1.2.08 (Paragrafo 3.4).

Il miglior assemblaggio viene stabilito basandosi sul numero di contigs o scaffolds prodotti, sulle dimensioni dei frammenti, sulla grandezza fisica del genoma (Mb) e sul valore di N50 e N90 (valore statistico volto a rappresentare che il 50/90 % dell'intero genoma è contenuto in scaffolds di dimensioni uguali o maggiori a quel valore). N50 è un parametro con un'attendibilità discutibile: possono essere generati genomi con valori molto elevati di N50 che ne possono ricoprire quasi l'intera ma all'interno dei contig possono essere presenti mismatch ed errori di assemblaggio. Affinché il valore di N50 sia affidabile è utile verificare l'assenza di errori nei contig assemblati (Schatz *et al.*, 2010).

Per *T. harzianum* T6776 l'assemblamento migliore (Kmer = 69) ha prodotto 1573 scaffolds, con contig di lunghezza massima di 33.10 Mb, N50 di 68.85 kb, per un totale di 39.73Mb. Il genoma di *T. gamsii* T6085, di dimensioni inferiori, è stato assemblato in 381 scaffold, con contig massimo di 11.99 Mb, N50 di 417.96 kb per un totale di 37.97 Mb (Kmer = 69).

I valori di assemblaggio sono stati comparati con i dati provenienti dai progetti di sequenziamento di altre specie di *Trichoderma* e con quelli dei diversi organismi scelti per l'analisi comparata (Tab. 4.2). È utile considerare che alcuni di questi organismi sono stati sequenziati con tecniche differenti, molto più accurate e costose di quella utilizzata nella presente tesi, pertanto possono essere presi come riferimento.

Le dimensioni dei genomi assemblati delle specie di *Trichoderma* si equivalgono. L'unico a mostrare un genoma più piccolo è *Trichoderma reseei* ma è anche l'unica specie ad essere saprotrofa tra le altre ad habitus micoparassitario. Le dimensioni inferiori del genoma di *Trichoderma reseei* 33.35 (Mb) potrebbero essere dovute alla perdita dei caratteri ancestrali del micoparassitismo (Kubicek *et al.*, 2011; Mukherjee *et al.*, 2012).

L'altra specie micoparassita presente nell'analisi è *C. rosea* che, come possiamo osservare nella Tabella 4.2, presenta una taglia del genoma molto elevata che rimane seconda solo a *Fusarium oxysporum*, e maggiore delle due specie oggetto di studio ma anche

dell'emibiotrofo (*C. fioriniae*), dei saprotrofi e dei patogeni. Questa differenza potrebbe essere dovuta alla produzione di un maggior numero di geni oppure alla presenza di porzioni ripetute più ampie.

Il numero di scaffolds prodotti nell'assemblaggio dei nostri genomi rientrano nella media dei genomi ottenuti con la stessa tecnologia di sequenziamento e di assemblamento, come in *C. fioriniae*, ma anche con il genoma di *C. rosea* ottenuto, sempre, con la tecnologia Illumina. Il genoma di *T. harzianum* T6776 presenta un numero di scaffolds maggiore rispetto al *T. gamsii* T6085 nonostante la taglia e la quantità di dati sia pressoché la stessa. Il motivo di tale frammentazione potrebbe risiedere nelle zone ripetute del genoma che sono difficoltose da assemblare.

La disponibilità di sequenze genomiche complete facilita l'annotazione e permette la costruzione di relazioni evolutive tra geni (Parra *et al.* 2007). La completezza dell'assemblamento è stata verificata sulla base del contenuto genico attraverso la pipe line "CEGMA v. 2.4". Nel genoma di *T. harzianum* sono stati individuati il 96.8% di geni interi e il 98.4% includendo i geni parziali, mentre nel genoma di *T. gamsii* sono presenti il 95.6% di geni interi e il 97.6% considerando anche i geni non completi. I risultati sono stati confrontati con quelli degli organismi scelti per le analisi comparate (Tab 4.2), osservando che i valori ottenuti per genomi completamente assemblati in cromosomi come *Neurospora crassa* e *Magnaporthe oryzae*, sono simili se non inferiori a quelli ottenuti per i due isolati di *Trichoderma*. Possiamo quindi approssimare che gli assemblaggi dei genomi oggetto di studio sono, da un punto di vista del contenuto genico, completi al 100%.

Code	Scaffolds	Assembly Size (Mb)	Max contig (bp)	N50 (bp)	N90 (bp)	CEGMA tot	CEGMA par
NCRA	20	41.04	9798.893	6000.761	4218.384	95.56%	97.18%
MORY	30	41.70	4429.722	2890.137	851.181	95.16%	97.58%
CFIO	1096	49.00	596.408	137.254	38.253	96.77%	98.39%
CPUR	191	32.09	977.986	433.221	110.809	94.76%	97.58%
TATR	29	36.14	5621.775	2007.903	1127.681	93.15%	95.16%
TGAM	381	37.97	1198.811	417.961	106.959	95.56%	97.58%
TREE	121	33.35	2198.954	837.556	224.034	95.56%	97.18%
THAR	1573	39.73	330.970	68.846	15.338	96.77%	98.39%
TVIR	93	39.02	3456.807	1836.662	328.624	95.16%	97.98%
FVER	36	41.78	4627.337	1959.799	825.785	95.97%	97.58%
FOXY	141	61.36	4358.238	1976.106	500.264	95.56%	97.18%
FGRA	31	36.45	8931.406	5350.016	2732.284	94.35%	97.18%
CROS	595	58.33	3780.011	794.244	82.291	91.53%	95.56%

Tabella 4.2. Valori statistici dei parametri di assemblamento di *T. harzianum*, *T. gamsii* (in grassetto) e degli organismi scelti per l'analisi comparata. Sono riportati il numero di scaffolds, la dimensione del genoma assemblato (Mb), il contig massimo, N50 e N90. Nelle ultime due colonne sulla destra sono presenti le percentuali (%) della completezza dell'assemblamento di CEGMA, totale e parziale.

4.4 Predizione genica

La predizione del contenuto genico dei genomi sequenziati ed assemblati è stata ottenuta mediante la pipe-line MAKER2. Sulla base del controllo automatico fornito durante il processo di invio dei genomi in banca dati (NCBI), sono stati sistemati manualmente eventuali errori, presenti nei geni, a carico dei siti di splicing. Sono stati predetti 11501 geni per *T. harzianum* e 10944 per *T. gamsii* (Tab.4.3).

La valutazione del contenuto in guanina e citosina nelle varie porzioni genomiche è un parametro utile in quanto fornisce indicazioni sul livello di stabilità di tali regioni oltre ad essere una caratteristica strutturale degli organismi. Il valore di GC medio riscontrato nei due genomi sequenziati (48.50 e 48.76%) è comparabile al valore medio presente negli organismi appartenenti al phylum Pezizomycotina (50.0-52.0%) (Amselem *et al.*, 2011). I genomi analizzati nelle regioni codificanti (esoni) manifestano un contenuto in guanina e citosina maggiore del 53.2% per entrambi gli organismi. Un valore minore di circa 9 punti percentuali si trova nelle zone non codificanti, come introni e zone intergeniche, con un punteggio leggermente inferiore nelle regioni introniche.

	<i>T. harzianum</i> T6776	<i>T. gamsii</i> T6085
Genome size (Mb)	39.73	37.97
GC (%)	48.50%	48.76%
Number of predicted genes	11501	10944
Total genes length (bp)	19466.93	18471.52
Genes GC (%)	52.3%	52.2%
Number of exons	32850	30611
Total exons length (bp)	17443.15	16543.06
Exons per gene (average)	2.86	2.8
Exons GC (%)	53.2%	53.2%
Numer of introns	21349	19667
Total introns length (bp)	2023.77	1928.47
Introns per gene (average)	1.86	1.8
Introns GC (%)	44.33%	43.85%
Intergenic regions size (Mb)	20.26	19.50
Intergenic regions GC (%)	44.81%	45.47%

Tabella 4.3. Valori statistici del contenuto genico predetto da MAKER2 pipeline per gli isolati *T.*

harzianum T6776 e *T. gamsii* T6085.

I due gli isolati mostrano il numero di geni, lunghezza totale dei geni (bp), contenuto medio di esoni e di introni per gene simile alle altre specie micoparassite di *Trichoderma* presenti nel lavoro di Kubicek *et al.*, 2011 e confrontabile anche con altri micoparassiti appartenenti a generi differenti come *C. rosea* (Karlsson *et al.*, 2015). Tali dati avvalorano il sistema di predizione genica utilizzato.

L'insieme dei prodotti genici predetti, detto anche proteoma, è stato utilizzato come base di partenza per l'analisi di comparazione genomica.

4.5 Analisi di comparazione genomica

Al fine di individuare le famiglie proteiche espanse o ridotte, coinvolte nell'adattamento evolutivo e nello stile di vita degli isolati T6776 e T6085, è stato effettuato un confronto con una serie di organismi modello con comportamenti ecologici differenti ma evolutivamente vicini al genere *Trichoderma* (Capitolo 3; tabella 3.1).

A partire dai proteomi dei vari organismi è stata individuata la porzione di proteine secrete. Nel grafico a blocchi (Fig. 4.5) sono riportate le dimensioni dei proteomi, dei secretomi e le percentuali delle putative proteine secrete per ogni genoma analizzato. Le dimensioni dei secretomi varia dal 10.0% al 16.0% del proteoma totale. Possiamo osservare che il numero di proteine secrete cambia in funzione dello stile di vita dell'organismo: gli organismi aventi lo stesso comportamento ecologico hanno anche porzioni di proteine secrete con valori simili.

Sulla base di tale osservazione possiamo dedurre che il contenuto di proteine secrete varia in relazione al grado di complessità delle interazioni che un organismo deve instaurare con il proprio ospite e/o con l'ambiente circostante per occupare una determinata nicchia ecologica.

Ad esempio *Claviceps purpurea*, biotrofo obbligato, ha un basso grado di complessità, poiché deve occupare una singola nicchia ecologica (rappresentata dall'ospite) ed ha il contenuto di proteine secrete (871, pari al 9.87% del proteoma totale) più basso in assoluto rispetto agli altri organismi analizzati. La taglia ristretta del secretoma rispecchia l'ecologia dell'organismo (che necessita di instaurare un'interazione esclusiva con la cellula ospite

per assicurare la propria sopravvivenza) e sottolinea la mancata interazione con l'ambiente. Al contrario, organismi emibiotrofi come *C. fioriniae* e *M. oryzae* hanno un secretoma molto espanso (rispettivamente pari al 16.01% e al 16.57%) a giustificare la maggiore complessità del proprio stile di vita. Infatti, durante il processo d'infezione della pianta questi patogeni passano da una fase iniziale biotrofa, tenendo in vita la cellula ospite, a una fase secondaria necrotrofa che porta alla morte di quest'ultima (O'Connel *et al.*, 2012). Infine, patogeni di tipo necrotrofo quali *F. oxysporum* e *F. graminearum*, mostrano un livello intermedio di proteine secrete (10.93% e 11.58%) correlato con il livello intermedio di complessità ecologica.

Un discorso a parte va fatto per i saprotrofi, cioè quegli organismi non patogeni per le piante. Nel nostro caso questo gruppo può essere suddiviso in due sottogruppi: il primo che include specie che non stabiliscono alcuna interazione con la pianta o con altri organismi (*T. reesei* e *N. crassa*) e il secondo che racchiude le altre specie di *Trichoderma* analizzate (Capitolo 3; Tabella 3.1) e *C. rosea*. Queste specie annoverano isolati noti per l'attività micoparassitaria e isolati in grado di colonizzare e promuovere lo sviluppo delle piante. *T. reesei* e *N. crassa* mostrano valori del secretoma piuttosto bassi e confrontabili con quelli dei biotrofi, pari circa al 10% del proteoma totale, che potrebbero essere giustificati da una minore complessità dell'ambiente che devo colonizzare. Al contrario, gli altri isolati di riferimento di *Trichoderma* e l'isolato di *C. rosea* mostrano percentuali più elevate (comprese tra 11.22% e 11.65%) mentre i nostri isolati raggiungono percentuali leggermente maggiori, pari a 12.28% per *T. harzianum* T6776 e 12.39% per *T. gamsii* T6085. Questi valori sono nettamente superiori a quelli dei biotrofi obbligati e dei patogeni necrotrofi ma decisamente inferiori a quelli degli emibiotrofi.

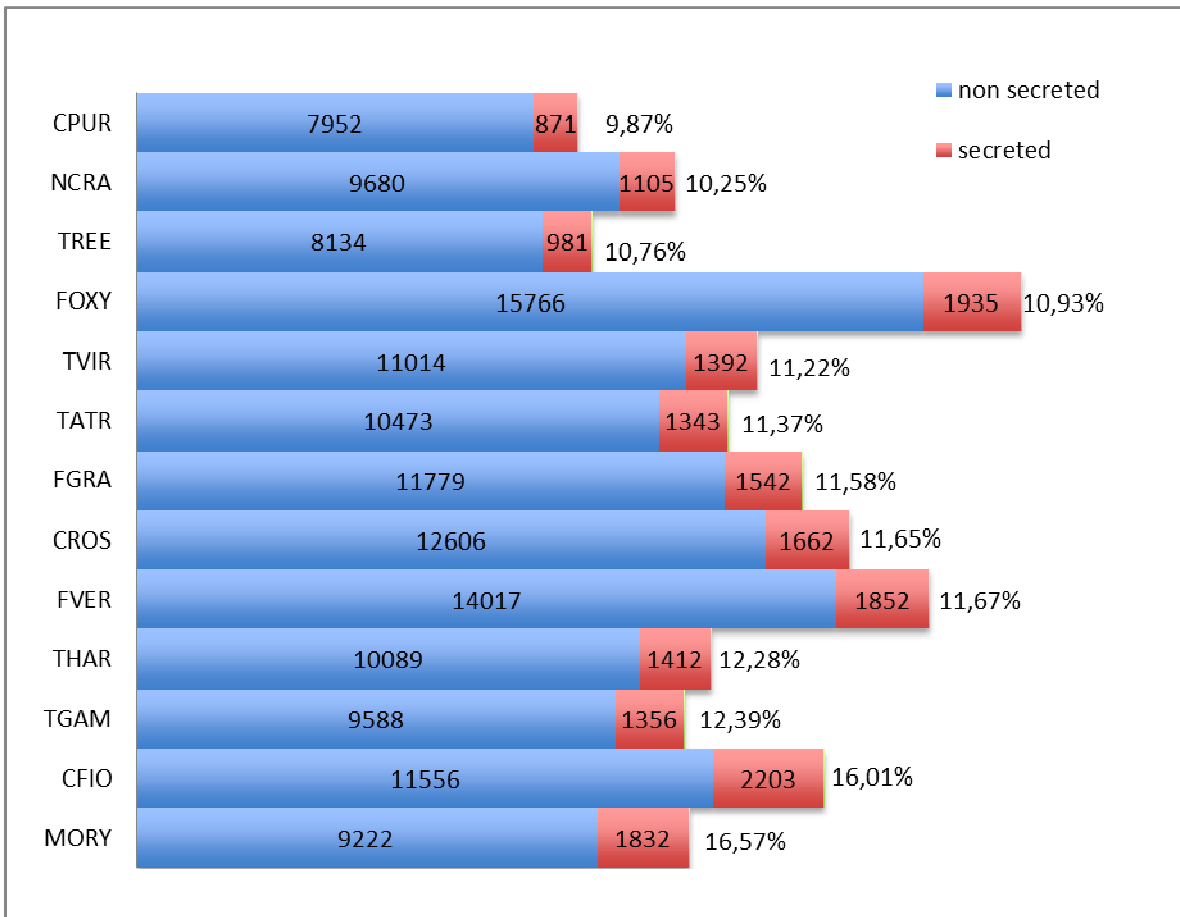


Figura 4.5. Grafico a blocchi della percentuale di proteine secrete (%) nei 13 genomi fungini analizzati. Per ogni organismo possiamo osservare il numero di proteine non secrete nella porzione in azzurro, numero di proteine secrete nella parte in rosso e nella colonna esterna all'immagine, la percentuale di proteine secrete rispetto al proteoma.

4.6 Famiglie geniche espanse in Trichoderma

Le proteine putative secrete sono state analizzate con InterProScan, con lo scopo di individuare le famiglie proteiche presenti nei genomi sulla base del riconoscimento dei domini o motivi proteici in esse conservati. Le funzioni proteiche sono state assegnate manualmente con l'ausilio del database InterPro.

La tabella 4.5 mostra le famiglie geniche maggiormente espanse negli organismi agenti di biocontrollo come *Trichoderma* e *Clonostachys*, ponendo un'attenzione particolare sugli isolati descritti nella presente tesi e confrontandole con gli altri organismi patogeni. Il livello di espansione è rappresentato dal numero di proteine secrete contenenti uno o più domini proteici associati ad una funzione biologica. In questo caso, il livello di espansione per ogni famiglia proteica all'interno dei diversi organismi presi in analisi viene osservato

in funzione del numero e del grado di intensità del colore della rispettiva casella presente nella tabella: maggiore è il livello di espansione più la colorazione è intensa. Le famiglie geniche con maggiore rilevanza sono state analizzate dividendole in base alla loro funzione; parte di esse appartengono a classi note e descritte in diversi organismi fungini, mentre altre hanno ancora un ruolo da esplorare.

4.6.1 Cell-wall-degrading enzymes (CWDEs)

All'interno della classe enzimatica coinvolta nella degradazione dei carboidrati e polisaccaridi presenti nella parete cellulare vegetale e fungina, vi sono le famiglie appartenenti ai CaZY (*Carbohydrate active enZYme*).

Le famiglie di CaZY espanso nelle specie micoparassite di *Trichoderma*, sia oggetto del presente studio che utilizzate per l'analisi comparata, sono le glicoside idrolasi GH5, GH30, GH13, GH2, GH20, GH2/20, le chitosanasi (GH75) e le polisaccaride liasi (PL7). Le glicoside idrolasi GH5 sembrano avere una funzione associata a una β -1,4-endoglucanasi con attività xyloglucanasi tipica dei patogeni delle piante, coinvolta nella degradazione delle emicellulose della parete vegetale (Kubicek *et al.*, 2014). Infatti rappresenta una famiglia molto espansa nei basidiomicota biotrofi come le ruggini del frumento quali *Puccinia graminis* e *Puccinia triticina* (Zhao *et al.*, 2013).. La produzione delle GH5 negli isolati sequenziati in questo lavoro potrebbe essere legata alla capacità di colonizzare le radici vegetali da parte di *T. harzianum* 6776 come endofita, e alla capacità di T6776 e T6085 di degradare i glucani presenti nel micelio di organismi patogeni come *R. solani*. Le GH30 (glucosilceramidasi) sono coinvolte nella scissione dei glicolipidi delle membrane cellulari in glucosio e ceramide. Entrambe le famiglie GH5 e GH30 sono state trovate anche in altre specie di *Trichoderma* con molteplici attività litiche quali mannosidasi, cellulasi e xyloglucanasi (Druzhinina I.S. *et al.*, 2012). La famiglia delle glicoside idrolasi GH13 contiene al suo interno un vastissimo numero di sottofamiglie aventi funzioni più svariate. In *Bacillus* spp. è stata studiata la funzione delle GH13 come α -amilasi e α -glucosidasi, che attraverso una reazione di transglicosilazione rompono le molecole di oligosaccaridi di amido in sei, sette o otto residui di glucosio rilasciandoli come ciclodestrine (Rye and Withers., 2000), molecole note anche come elicitatori per la pianta. La produzione di questa famiglia enzimatica da parte di *Trichoderma* potrebbe quindi far pensare ad una possibile funzione come elicitatori per il riconoscimento nell'interazione pianta- fungo. Le famiglie di glicoside idrolasi GH2, GH20, GH2/20 e le

β -esosaminidasi (IPR029019) sono chitinasi coinvolte nell'idrolisi della chitina, polimero costituente principale della parete cellulare fungina. Precisamente le GH2 sono β -esosaminidasi che tagliano i legami beta della catene di chitina, mentre le GH20 sono chitobiasi che attaccano una porzione zuccherina proveniente dalla degradazione della chitina, il chitobiosio. Un'altra classe di enzimi coinvolti nella degradazione del chitosano delle ife e del micelio fungino sono le chitosanasi (IPR009939) appartenenti alla famiglia delle GH75. Nel lavoro di Kubicek *et al.*, 2011 incentrato sull'origine ancestrale della capacità micoparassitaria di *Trichoderma*, questa famiglia è presente in *T. virens* e in *T. atroviride*.

Oltre alle suddette famiglie di glicoside idrolasi, sono presenti anche le polisaccaride liasi della famiglia 7, precisamente le alginato liasi (IPR008929). Le alginato liasi catalizzano la depolimerizzazione dell'acido alginico o alginato, polimero dell'acido D-mannuronico. Sono implicate nel metabolismo dei polimeri del mannosio, come i mannani presenti nella parete cellulare fungina. Questa funzione potrebbe spiegare l'espansione delle alginato liasi negli isolati di *Trichoderma* micoparassiti e in *C. rosea*. Inoltre, la produzione delle alginato liasi negli isolati *T. harzianum* T6776 e *T. gamsii* T6085 è un'informazione molto importante al fine di una possibile applicazione dei due isolati come biofitofarmaci in preparati commerciali dove l'alginato di calcio e l'alginato di sodio vengono impiegati come adesivanti, andando a favorire la rottura dei legami tra le molecole di alginato per il rilascio del principio attivo. Studi precedenti hanno valutato la capacità alginolitica, presumibilmente dovuta alla presenza delle alginato liasi con attività endolitica, in diversi Ascomycota, compresi isolati appartenenti al genere *Trichoderma* fatti crescere su substrati contenenti alginato come unica fonte di carbonio. Nonostante la completa crescita dei miceli sul substrato contenente alginato, non sono state registrate attività riconducibili alle alginato liasi, misurata analizzando la viscosità del substrato, in isolati di *Trichoderma* (Sarrocco *et al.*, 2004).

Le glucanasi sono un gruppo di enzimi litici della parete cellulare, aventi un ruolo all'interno del micoparassitismo e biocontrollo in *Trichoderma* (Mukherjee *et al.* 2012). I glucani sono polisaccaridi lineari formati da unità di glucosio che vanno a legare i polimeri di chitina o chitosano all'interno della parete cellulare fungina. Vi sono due tipi di glucani che si differenziano per il tipo di legame chimico stabilito tra le unità di glucosio: i beta-glucani, precisamente i β -1,3 e i β -1,6 glucani, che danno rigidità alla parete cellulare e gli alfa-glucani caratterizzati dagli α -1,3 e α -1,4 che vanno a costituire la matrice cellulare (Daguerre Y. *et al.*, 2014). Le glucanasi a loro volta si dividono in eso e endo glucanasi e la

loro funzione è quella di idrolizzare i residui di glucosio presente nei β -glucani (Lopez-Mondejar *et al.*, 2010). Sono state trovate numerose putative glucanasi aventi un ruolo all'interno del micoparassitismo. Le β -1,3 glucanasi come la Glu78 è stata trovata in *T. atroviride* con attività antifungina, ne sono state identificate altri tipi anche in *T. harzianum*, *C. rosea* che hanno funzione litica nei confronti di patogeni o di antagonisti (Daguerre *et al.*, 2014). Il coinvolgimento delle β -1,3 exoglucanasi è stato studiato in *T. harzianum* (Kupicek R *et al.*, 1999). Nel nostro caso, la presenza di questa famiglia è stata osservata in tutti gli organismi selezionati ma in misura maggiore nei genomi sequenziati di *T. harzianum* e *T. gamsii*, in *C. rosea* e in *C. fioriniae* associata all'espansione del dominio proteico carbohydrate-binding WSC (IPR013994).

Un'attenzione particolare è stata posta sul dominio della famiglia proteica EXTL2 (IPR015338), una glucosil-transferasi della famiglia 47 coinvolta nel trasferimento dei polimeri di N-acetilglucosammina a molecole trasportatrici dello zucchero come UDP-glucosio, all'interno del processo biosintetico dei polimeri della parete cellulare fungina. La presenza di questa famiglia è stata osservata nelle specie micoparassite, in *T. reesei* e completamente assente negli altri organismi.

4.6.2 Peptidasi

Le peptidasi sono enzimi proteolitici prodotti e secreti da tutti gli organismi fungini, compresi i ceppi agenti di biocontrollo di *Trichoderma*, in quanto giocano un ruolo significativo nell'antagonismo. Simili ad altri sistemi litici, le proteasi partecipano alla degradazione delle proteine cellulari durante l'attacco parassitario ai loro ospiti, in modo da destabilizzare l'integrità delle membrane cellulari e facilitarne la penetrazione e la colonizzazione (Suárez *et al.*, 2007). Dalle analisi condotte, le aspartato peptidasi, le peptidasi G1 e le serine peptidasi S53 sono state trovate maggiormente espresse in *T. harzianum* T6776 e in *T. gamsii* T6085.

Le aspartato peptidasi, secrete da tutti gli organismi presi in analisi, ma in misura leggermente maggiore in *T. harzianum* T6776, sono state ampiamente studiate in diverse specie appartenenti a questo genere mostrando un ruolo nella scissione della parete cellulare di molti fitopatogeni oppure disattivando le endo- e eso-poligalatturonasi prodotte da patogeni come *B. cinerea* (Dou *et al.*, 2013). In alcuni isolati di *Trichoderma*, queste proteine agiscono degradando e disattivando i recettori coinvolti nella difesa delle

piante in modo da facilitarne la colonizzazione e risultando in una serie di effetti benefici, quali promozione della crescita e difesa da patogeni, in particolare da quelli tellurici (Viterbo *et al.*, 2004). Data la particolare attitudine quale endofita, quest'ultimo ruolo descritto potrebbe essere associato all'espansione di questa classe in *T. harzianum* T6776. Le serine peptidasi S53, dette anche sedolisine o carbossi peptidasi, sono particolarmente espresse nei due isolati di *Trichoderma* oggetto di studio. Svolgono un'azione antifungina, inibendo la crescita del micelio di molti funghi fitopatogeni e, in *T. pseudokoningii*, un'attività nematocida che porta alla morte delle uova e delle larve del nematode parassita delle piante *Meloidogyne incognita* (Druzhinina *et al.*, 2012). Altre classi di serine peptidasi come la S8, oltre che nel genere *Trichoderma*, sono state studiate anche nei biotrofi endofiti con un ruolo nella degradazione dei recettori proteici di membrana presenti nelle piante come segnali di difesa. Infine, le glutammato peptidasi G1 (Aspergillo pepsine o scitalidoglutammato peptidasi) prodotte in Pezizomycotina come *B. cinerea* (Amselem *et al.*, 2011) ed in Sordariomycotina, sono coinvolte nella degradazione dei polimeri della parete durante il processo d'infezione da parte dei patogeni e la loro azione dipende dal pH acido della cellula vegetale. Inoltre, l'inibitore proteasico (IPR009020) è particolarmente espanso negli isolati oggetto di studio e nel micoparassita *C. rosea* ed è coinvolto nell'induzione o repressione dell'attività delle serine peptidasi S53 e S8.

4.6.3 Proteine coinvolte nella sintesi di metaboliti secondari

Dalle analisi condotte all'interno del pool di proteine secrete dagli isolati di *Trichoderma* oggetto di studio, sono emerse anche proteine coinvolte nella biosintesi di polichetidi o peptaboliti, composti legati alla produzione di metaboliti secondari durante i meccanismi di antibiosi fungina.

In generale, questi prodotti genici non vengono secreti al di fuori del compartimento cellulare fungino, in quanto sono coinvolti nel metabolismo delle cellule che li sintetizzano. Questo fa supporre che nelle specie di *Trichoderma* queste proteine siano secrete proprio per poter interferire con il metabolismo secondario della pianta.

Appartenenti a questa classe proteica vi sono le famiglie di polichetide sintasi espresse nelle specie di *Trichoderma*, come le polichetide ciclasi/deidratasi (IPR019587), le polichetide sintasi acidi grassi sintasi (IPR020842) e le polichetide sintasi enoilreduttasi (IPR020843).

Le polichetide ciclastasi deidratasi sono prodotte esclusivamente dal genere *Trichoderma* e assenti in tutti gli altri organismi. La loro funzione sembra essere associata ai meccanismi di trasduzione del segnale eucariotici legati ai legami lipidici e agli antigeni delle proteine vegetali prodotte in risposta agli stress e agli attacchi di patogeni. Potremmo ipotizzare che la produzione di questa famiglia enzimatica, in *Trichoderma*, abbia la funzione di agire come repressore dei segnali di difesa della pianta per incentivarne la colonizzazione.

Sulla possibile funzione delle altre due famiglie di polichetide sintasi all'interno degli isolati di *Trichoderma* non sono state trovate molte informazioni, soprattutto riguardanti il sistema descritto. Per tale motivo, questo dato potrebbe essere un punto di partenza per ulteriori studi di ricerca futuri.

4.6.4 Famiglie proteiche non appartenenti a classi specifiche

Sono state trovate famiglie particolarmente interessanti come le epossido idrolasi (IPR016292) il cui ruolo nel micoparassitismo non è ben definito ma che si sa essere implicate nella degradazione di composti organici. Un'altra famiglia interessante è quella associata al dominio lisozima-like (IPR023346) che sembra svolgere un ruolo antibatterico. Una maggiore espansione in *T. harzianum*, *T. gamsii* e *C. rosea* è stata individuata per le idrolasi con funzione di esterasi appartenenti alla famiglia Alpha/beta idrolasi fold-1 (IPR000073). Inoltre, è presente anche il dominio associato alla famiglia delle idrofobine cerato-ulmine (IPR010636). La cerato-ulmina è stata isolata per la prima volta in *Ophiostoma ulmi* dove è stato visto essere coinvolta nei meccanismi di patogenicità (Temple *et al.*, 1997). Le idrofobine sono piccole molecole di circa 70- 400 aminoacidi, con un dominio formato da otto residui di cisteina in posizione conservata tenute insieme da legami bisolfito (Degenkolb *et al.*, 2008). Queste proteine ubiquitarie presenti nei funghi filamentosi svolgono molteplici ruoli all'interno di specie diverse ma anche all'interno della singola specie.

In *Trichoderma* sono coinvolte nella formazione di ife aeree (Viterbo & Chet., 2006) e hanno un ruolo come molecole segnale; vengono prodotte a livello della superficie della parete ifale o dei conidi durante la prima fase del micoparassitismo quando c'è contatto con l'ospite inducendo la formazione di coilings (Mukherjee *et al.* 2013). In *Trichoderma reesei* sono state caratterizzate due classi di idrofobine: HFBI legate alla parete cellulare con un ruolo nello sviluppo delle ife, mentre HFBII vengono secrete come proteine extracellulari

inducendo la sporulazione (Viterbo and Chet., 2006). E' nota la presenza di 6 geni codificanti le idrofobine in *T. reesei*, 9 in *T. virens* e 10 in *T. atroviride* appartenenti alla classe HFBII (Lorito *et al.*, 2010). In *T. asperellum* e in *T. harzianum* è stata dimostrata la funzione delle idrofobine (TasHyd1 e Qid74) nel migliorare l'adesione, l'ancoraggio, la colonizzazione dell'apparato radicale delle specie vegetali (Mukherjee *et al.* 2013). La stessa funzione svolta dalle idrofobine nei meccanismi di biocontrollo e colonizzazione delle radici è stata scoperta in *C. rosea* (Daguerre Y. *et al.*, 2014). Inoltre, l'aumento o la diminuzione del contenuto genico di idrofobine all'interno del genoma potrebbe avere un ruolo nel processo di speciazione (Degenkolb *et al.*, 2008).

Dalle analisi condotte sui genomi di *T. harzianum* T6776 e *T. gamsii* T6085 è stato osservato il gene codificante per la famiglia proteica WD40 repeat (IPR001680). Si tratta di proteine costituite da motivi di 40 amino acidi ripetuti, coinvolte nei meccanismi di traduzione del segnale e regolazione trascrizionale nelle cellule eucarioti. Queste proteine sono particolarmente espanse in *T. gamsii* e potrebbero agire interferendo sui sistemi di regolazione vegetali. Sempre in *T. gamsii* è prodotta in maggior misura la famiglia proteica NmrA-like (IPR008030), un fattore di regolazione negativo delle proteine NmrA che sembrano essere coinvolte nel metabolismo dell'azoto. Un'ipotesi sulla funzione di questa proteina legata al metabolismo dell'azoto potrebbe essere fatta in funzione dello studio che rivela che la produzione di proteasi durante la prima fase del micoparassitismo sembrerebbe essere indotta a causa della limitazione della disponibilità dell'azoto in *T. atroviride* (Seidl *et al.*, 2009).

Sebbene non espanso nei due isolati sequenziati, è presente anche un dominio isocorismatasi-like (IPR000868). L'isocorismatasi è un enzima che catalizza la conversione da isocorismato a 2,3-diidrossibenzoato e piruvato. È implicato nella sintesi del composto antimicrobico fenazina da *Pseudomonas aeruginosa*, *Enterobactina* e da *Escherichia coli* (Soanes *et al.*, 2007). Il motivo isocorismatasi è stato trovato anche in un certo numero di idrolasi, come le nicotinamidasi che convertono la nicotinammide in acido nicotinico. Inoltre, è un precursore dell'acido salicilico che viene sintetizzato nelle piante in risposta agli attacchi dei patogeni. È possibile ipotizzare che l'isocorismatasi secreta da molti ascomiceti filamentosi, patogeni e non, potrebbe agire come inibitore della risposta delle piante agli attacchi dei patogeni in modo da rendere le piante più suscettibili (Soanes *et al.*, 2008). Curiosa è la presenza del gene per le gliossalasi (IPR025870), in quanto sembrano essere coinvolte in meccanismi di detossificazione.

Le espansioni geniche osservate correlate al processo di micoparassitismo degli isolati di *T.*

harzianum T6776 e *T. gamsii* T6085 sono frutto di analisi preliminari, affinché siano confermati i dati ottenuti c'è bisogno di ulteriori analisi funzionali e di espressione.

InterPro Term	TATR	TGAM	TREE	THAR	TVIR	FVER	FOXY	FGRA	CROS	NCRA	MORY	CFIO	CPUR
IPR000250 Peptidase G1	6	5	3	5	2	1	1	1	1	2	2	1	1
IPR001969 Aspartic peptidase	11	11	10	14	11	11	9	9	8	9	9	10	6
IPR009020 Proteinase inhibitor	19	21	7	19	15	16	15	15	27	8	12	15	9
IPR015366 Serine peptidase S53	12	15	3	12	7	5	3	3	2	2	4	3	5
IPR001139 Glycoside hydrolase, family 30	5	5	5	6	6	2	3	0	2	2	0	2	1
IPR003159 Polysaccharide lyase family 8	1	1	1	1	1	0	0	0	0	0	0	0	0
IPR006102 Glycoside hydrolase, family 2	6	6	4	5	5	2	1	2	2	2	4	2	1
IPR009939 Fungal chitosanase	5	5	3	5	5	2	2	1	5	1	0	2	1
IPR013780 Glycosyl hydrolase, family 13	15	15	13	15	16	7	8	4	6	4	6	10	2
IPR013994 Carbohydrate-binding WSC	5	13	4	12	5	4	1	4	11	5	4	16	3
IPR014718 Glycoside hydrolase-type carbohydrate-binding	9	10	9	7	9	3	3	2	6	5	7	10	4
IPR013812 Glycoside hydrolase, family 2/20	6	6	4	5	5	2	1	2	2	2	4	2	1
IPR018087 Glycoside hydrolase, family 5	4	4	2	5	3	1	1	0	1	0	1	1	0
IPR025975 Polysaccharide lyase	2	2	2	2	2	0	0	1	2	1	1	1	0
IPR029019 Beta-hexosaminidase	2	2	2	2	2	1	1	1	1	1	1	1	1
IPR000073 Alpha/beta hydrolase fold-1	10	15	7	13	10	6	6	4	8	3	6	4	1
IPR000639 Epoxide hydrolase-like	8	15	4	10	7	3	6	2	6	0	1	3	0
IPR000868 Isochorismatase-like	4	3	1	2	3	1	2	1	5	0	2	1	0
IPR001568 Ribonuclease T2-like	4	4	3	4	5	2	3	3	4	1	2	2	2
IPR001680 WD40 repeat	8	4	14	8	11	3	7	5	12	4	5	9	4
IPR002347 Glucose/ribitol dehydrogenase	17	25	10	20	13	14	23	10	28	7	13	17	3
IPR008030 NmrA-like	9	13	7	8	7	1	4	3	10	0	0	3	1
IPR008929 Chondroitin AC/alginate lyase	4	4	2	4	4	1	2	1	5	0	0	3	0
IPR010497 Epoxide hydrolase	3	6	1	5	2	1	2	1	6	0	0	1	0
IPR010636 Cerato-ulmin hydrophobin family	9	9	6	5	8	3	3	1	2	1	3	3	4
IPR013785 Aldolase-type TIM barrel	16	14	12	14	14	8	11	5	12	0	5	8	2
IPR015338 EXTL2	1	1	1	1	1	0	0	0	2	0	0	0	0
IPR016292 Epoxide hydrolase	3	5	1	5	2	1	0	0	5	0	0	0	0
IPR018017 Nucleoside phosphorylase	6	7	1	2	8	1	5	2	2	0	2	0	0
IPR018188 Ribonuclease T2	3	3	3	4	4	2	3	2	4	1	2	1	2
IPR019587 Polyketide cyclase/dehydrase	1	1	1	1	1	0	0	0	0	0	0	0	0

IPR020842 Polyketide synthase/Fatty acid synthase	4	4	1	5	1	0	0	0	4	0	1	1	0
IPR020843 Polyketide synthase, enoylreductase	3	3	0	4	0	0	1	0	4	0	0	3	0
IPR023346 Lysozyme-like domain	5	6	4	7	5	1	1	0	6	0	1	1	2
IPR025870 Glyoxalase-like domain	2	2	0	1	0	1	1	0	3	0	1	0	0

Tabella 4.5. Nella heat map sono riportate le famiglie proteiche coinvolte nel processo di micoparassitismo del genere *Trichoderma*. Nella prima colonna a sinistra sono presenti i codici e i nomi delle famiglie proteiche, le colorazioni stanno ad indicare le classi enzimatiche di appartenenza: CaZy (giallo), peptidasi (verde), polichetide sintasi (arancione) e nessun colore per le classi non specifiche. Nelle colonne successive è indicato il numero di volte in cui una famiglia è presente nel genoma di un organismo. L'intensità della colorazione indica il livello di espansione.

4.7 Caratterizzazione dei CaZY (Carbohydrate active enZYme)

I CaZY sono stati determinati con dbCAN e annotati manualmente attraverso il database CaZY (www.cazy.org). I risultati degli isolati T6776 e T6085 sono stati comparati con i modelli fungini aventi stili di vita differenti. Il numero di enzimi di ogni organismo per ogni classe di CaZY è riportato nella Figura 4.6.

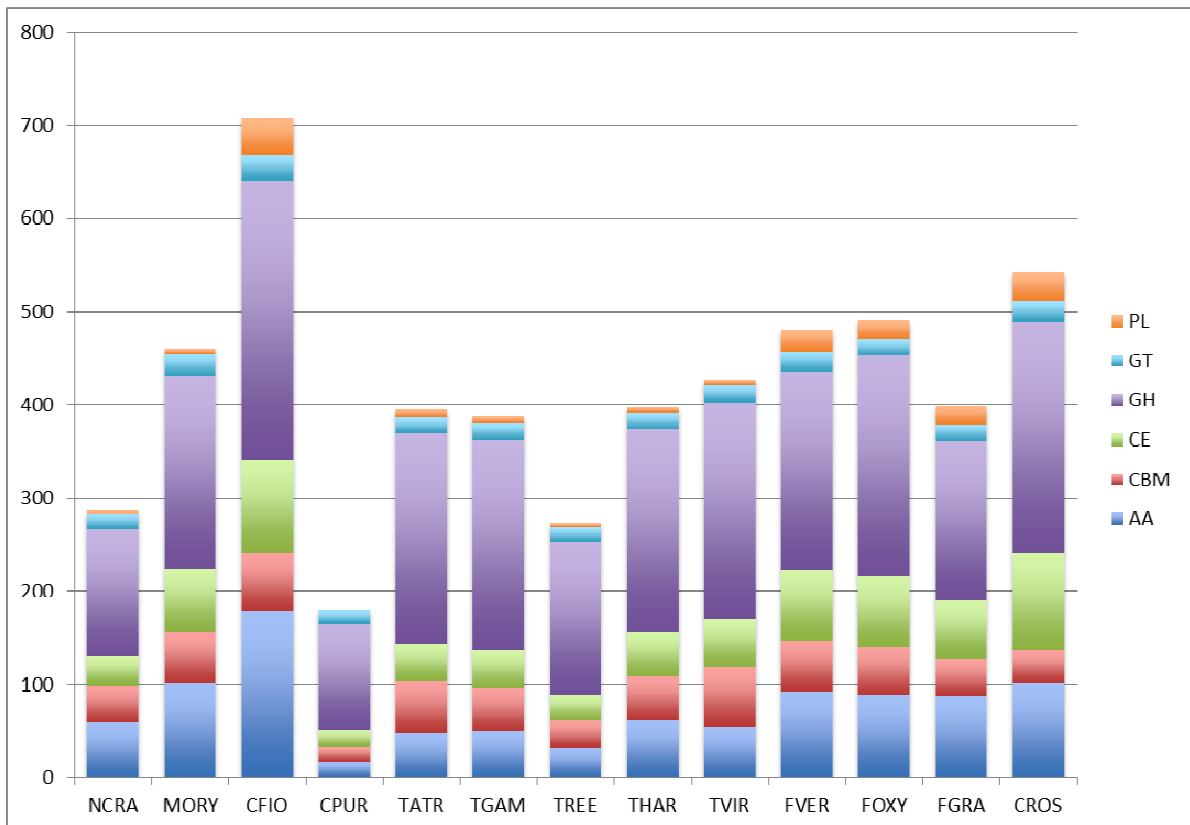


Figura 4.6. Diagramma a colonne dove sono rappresentate le espansioni delle classi geniche dei Carbohydrate active enzyme nei 13 genomi fungini. Sull'asse y è riportato il numero delle proteine predette, mentre sull'asse x vi sono i vari organismi. Le classi funzionali sono: polisaccaride liasi (PL), glicosil transferasi (GT), glicoside idrolasi (GH), Carbohydrate esterasi (CE), Carbohydrate binding module (CBM) e Auxiliary Activities (AA). Le colorazioni corrispondono a specifiche classi funzionali come riportato nella legenda.

Possiamo osservare che l'andamento dell'espansione genica in *Trichoderma* è simile in tutte le specie al di fuori dell'unica specie saprotrofa *T. reesei*. Al contrario delle previsioni attese, *C. rosea* ha mostrato un'espansione maggiore rispetto agli altri micoparassiti, soprattutto a carico delle carboidrati esterasi (CE) e degli enzimi ad attività ausiliaria (AA), mentre risulta più scarsa nella produzione dei CBM rispetto a *T. harzianum* e *T. gamsii*.

Gli isolati oggetto di studio hanno mostrato, approssimativamente, la stessa presenza genica in ogni classe enzimatica (Tab 4.6)

	AA	CBM	CE	GH	GT	PL	TOT
THAR	61	47	48	217	18	6	397
TGAM	50	46	41	225	18	8	388

Tabella 4.6. Valori delle espansioni delle classi geniche dei Carbohydrate active enzyme di *T. harzianum* T6776 e *T.gamsii* T6085.

Successivamente sono state identificate manualmente le famiglie di CaZY maggiormente espansive negli organismi con capacità di agenti di biocontrollo.

Le classi implicate sembrano essere: chitosanasi (GH75), chitinasi (GH15), le β -1-3-glucanasi (GH55 e GH64), alginato liasi (PL7) e α -L-fucosidase (GH95) (Fig. 4.7).

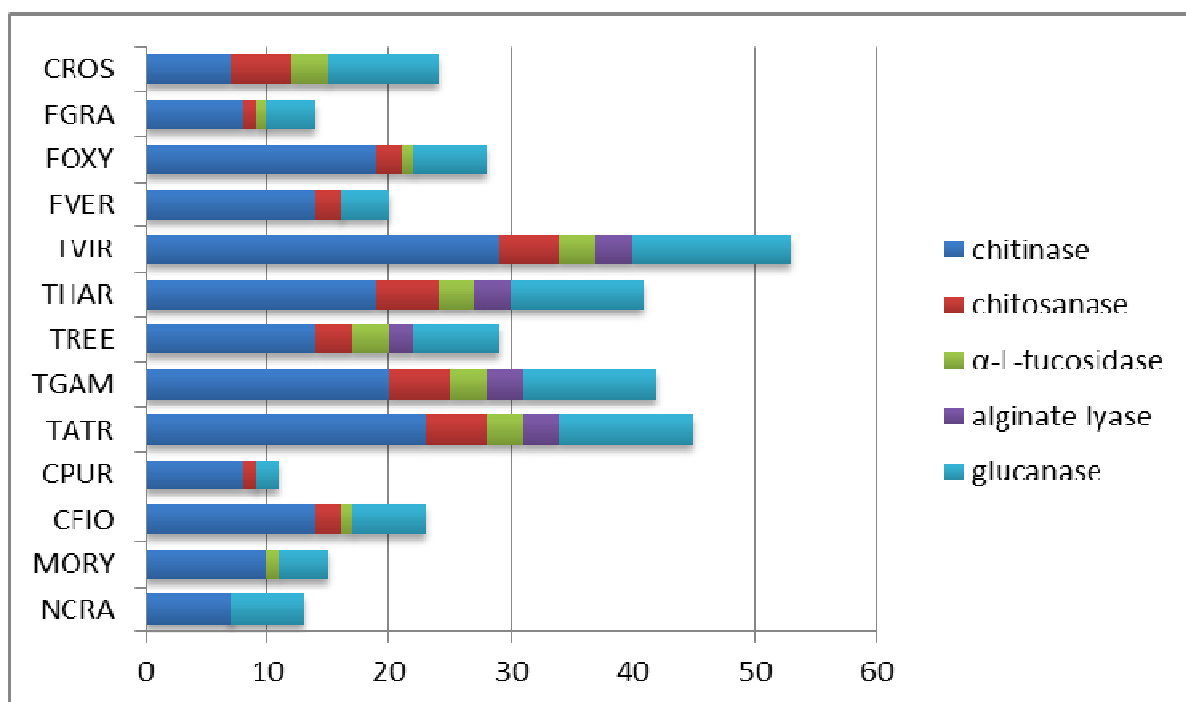


Figura 4.7 Grafico a blocchi dove vengono mostrate le espansioni geniche delle classi dei Carbohydrate active enzyme nei 13 genomi fungini implicati nel processo di micoparassitismo. Sull'asse X sono presenti i valori corrispondenti all'espansione; sull'asse Y sono presenti i genomi fungini. I colori dei blocchi corrispondono ad una specifica classe enzimatica come riportato nella legenda.

Alcune di queste, come le chitosanasi (GH75) e le β ,1-3 glucanasi (GH55 e GH64), vanno a confermare le classi coinvolte nel processo di micoparassitismo di *T. vires* e *T.*

atroviride descritte nel lavoro di Kubricek *et al.* (2011).

La presenza delle chitosanasi è molto interessante, in quanto il chitosano è un derivato parzialmente deacetilato della chitina e la presenza di questo composto nella parete cellulare fungina dipende dalla specie e dalle condizioni di crescita, inoltre in quantità maggiore si trova nelle ife mature. Studi riportano che un possibile meccanismo di difesa è quello di deacetilare la chitina (Kubricek *et al.* 2011).

L'annotazione dei CaZy conferma i risultati ottenuti dalle analisi IPR, dimostrando che questo tipo di analisi preliminare è un valido strumento per l'identificazione delle espansioni geniche nei genomi fungini. Con questo metodo non solo abbiamo confermato quanto già noto, ma siamo riusciti ad identificare anche nuove famiglie geniche espanse in questo gruppo tassonomico, se non specifiche dei due isolati oggetto di studio. Inoltre, come possiamo osservare nel grafico (Fig. 4.6), le espansioni dei CaZy secreti dagli organismi rispecchiano le percentuali ottenute nei secretomi (Fig. 4.5), dove gli emibiotrofi insieme ai patogeni si mostrano i maggiori produttori di proteine secrete e quindi anche di CaZy. Subito dopo troviamo i micoparassiti, specialmente *C. rosea* ed infine, i biotrofi con il contenuto più basso.

Questi dati confermano la correlazione esistente tra lo stile di vita o nicchia ecologica occupata da un organismo e la produzione delle classi enzimatiche coinvolte nell'adattamento ad essa.

5. Conclusioni

Negli ultimi 50 anni l'impiego dei mezzi chimici in agricoltura per il controllo della malattie e il mantenimento della fertilità del suolo ha giocato un ruolo chiave nel soddisfare la richiesta di incremento della produzione delle risorse alimentari nel mondo. La popolazione mondiale è in continuo aumento e a metà di questo secolo si stima raggiungerà i 9 miliardi di persone. La richiesta di cibo e risorse alimentari continuerà ad aumentare ma allo stesso tempo bisognerà far fronte alle esigenze ambientali che richiedono un'agricoltura più sostenibile e una diminuzione dell'impatto di quest'ultima sull'ambiente.

L'integrated pest management (IPM) è un approccio mirato alla riduzione dell'impiego di pesticidi in agricoltura attraverso l'uso di: pratiche colturali preventive, cultivar resistenti alle malattie e mezzi meccanici e biologici per il controllo della popolazione dei patogeni. La strategia di impiegare i microrganismi antagonisti e agenti di biocontrollo (BCAs) per ridurre la presenza dei patogeni in agricoltura, va a soddisfare entrambe le richieste del momento.

Trichoderma rappresenta un sistema modello sia per studi che riguardano i meccanismi relativi al biocontrollo (da anni sono in commercio formulati a base di *Trichoderma* impiegati per la protezione delle colture) sia per studi genetici legati ai diversi stili di vita di questi organismi (micoparassitario nei confronti dei patogeni e/o endofitico cioè capace di instaurare relazioni benefiche con le piante).

La necessità di sequenziare i due isolati di *Trichoderma*, oggetto di studio della presente tesi, con le note capacità sopra descritte nasce dall'esigenza di approfondire le conoscenze genetiche coinvolte in tali meccanismi caratterizzandone le principali famiglie geniche, ma soprattutto per creare una base solida per lo sviluppo di studi di genomica funzionale come trascrittomica, metabolomica e proteomica.

I genomi dei due isolati sono stati sequenziati ed assemblati *de novo*, la predizione genica è avvenuta in modo automatico e l'annotazione del contenuto genico è stata eseguita manualmente con l'ausilio di database specifici disponibili on-line. Il contenuto proteico dei due isolati è stato comparato con quello di altre specie di *Trichoderma* di cui sono stati resi accessibili i dati genomici e con quelli di organismi modello rappresentanti i vari stili di vita fungini: biotrofi, emibiotrofi, necrotrofi, saprotrofi e micoparassiti.

L'analisi comparata è stata eseguita mettendo a confronto la porzione di proteine secrete

dei vari organismi: i secretomi rappresentano infatti la totalità delle proteine extracellulari secrete da un organismo nell'ambiente esterno, aventi composizione e funzione specifiche per il ruolo ecologico ricoperto dall'organismo stesso.

Sulla base dei secretomi sono state determinate le espansioni e le riduzioni delle famiglie proteiche secrete, focalizzandoci in misura maggiore su quelle espansive riconducibili al processo di micoparassitismo dei due isolati di *Trichoderma* oggetto della presente tesi.

Nell'analisi sono state individuate varie classi proteiche note come CaZY (*Carbohydrate active enZYme*), coinvolte nella degradazione dei polisaccaridi presenti nella parete cellulare fungina e vegetale, le peptidasi coinvolte nella degradazione della porzione proteica presente nelle membrane cellulari, oltre che famiglie con funzioni ancora da determinare che potrebbero rappresentare il target per progetti futuri.

Le dimensioni dei secretomi ottenuti per i singoli organismi rispecchiano il grado di complessità delle interazioni caratteristiche dei vari stili di vita, confermando i biotrofi e i necrotrofi con il contenuto di proteine secrete più ristretto, gli emibiotrofi con le dimensioni più elevate e i micoparassiti in linea con le dimensioni dei patogeni.

Nell'ambito delle famiglie proteiche dei CaZy, le famiglie maggiormente espansive sono state dettagliatamente studiate nei nostri isolati e confrontate con quelle degli organismi scelti come riferimento per i diversi stili di vita. I risultati ottenuti hanno mostrato una diversa complessità di questa classe enzimatica, in linea con la complessità ecologica degli stili di vita degli organismi studiati.

Gli isolati di *Trichoderma* T6776 e T6085 si sono rivelati dei forti produttori di chitinasi, chitosanasi e β ,1-3 glucanasi, enzimi coinvolti nelle modifiche e nella degradazione della chitina nella parete fungina durante il processo di micoparassitismo. Al contempo non sono state osservate forti espansioni delle cutinasi e pectinasi (come le pectino liasi, pectato liasi e poligalatturonasi), enzimi coinvolti nella degradazione della parete delle cellule vegetali o prodotte durante le interazioni fungo-pianta, legate più ad uno stile di vita endofitico e normalmente presenti in numero elevato in diversi patogeni. Questo potrebbe essere dovuto al fatto che entrambi gli isolati di *Trichoderma* appartengono a specie in cui il carattere micoparassitario ha origine ancestrale e che, solo successivamente, hanno sviluppato uno stile di vita endofitico.

Questo può essere riferito soprattutto all'isolato di *T. harzianum* T6776 noto per la capacità endofitica verso molteplici specie vegetali, per l'attività di promozione della crescita degli organi vegetali e per gli effetti benefici (biochimici) durante l'interazione con la pianta. Studi precedenti su *T. gamsii* T6085 hanno evidenziato le sue abilità come

micoparassita ad esempio nei confronti di specie micotossigene e fitopatogene di *Fusarium* piuttosto che come endofita di specie vegetali. I risultati relativi al contenuto genico analizzato vanno a confermare quanto già si sapeva circa i due diversi stili di vita.

Particolarmente interessanti sono risultate le espansioni geniche caratteristiche del genere *Trichoderma* come: le epossido idrolasi che giocano un ruolo interessante in un possibile impiego nella degradazione di composti organici oppure le alginato liasi che potrebbero essere impiegate nel rilascio del principio attivo incluso in biofitofarmaci impieganti alginato di calcio e alginato di sodio come coformulanti.

Infine, di particolare interesse è la famiglia delle polichetide sintasi e delle polichetide ciclasi deidratasi. Proteine appartenenti a questa famiglia non dovrebbero essere presenti nel secretoma, in quanto le loro funzioni vengono esplicate all'interno della membrana cellulare fungina per la biosintesi di metaboliti secondari che vengono secreti extracellularmente. La presenza di così tante proteine prodotte da geni coinvolti nel metabolismo secondario, lascia ipotizzare un ruolo nella modulazione o interferenza con il metabolismo secondario dell'ospite forse per inibirne i sistemi di difesa e favorirne la colonizzazione.

Ulteriori studi di genomica funzionale (come trascrittomica, metabolomica, proteomica, ecc...) sono necessari per convalidare i dati ottenuti dall'analisi preliminare sin qui svolta, che ci ha permesso di identificare le famiglie geniche target dello stile di vita dei due isolati di *Trichoderma* oggetto della presente tesi, confermando l'analisi comparata come un potente mezzo d'indagine.

I genomi sequenziati e assemblati degli isolati *T. harzianum* T6776 e *T. gamsii* T6085 e i dati relativi alle espansioni delle famiglie geniche coinvolte nei meccanismi di micoparassitismo e biocontrollo, non solo potranno essere utilizzati come piattaforma per altri studi di genomica comparata ma anche per altri lavori omici come:

- *Analisi trascrittomica*: si potrebbe impiegare lo stesso isolato fungino in due sistemi differenti: un sistema riguardante l'interazione pianta-fungo per osservare le porzioni di genoma coinvolte nell'endofitismo e nei meccanismi di induzione legati all'effetto benefico di *Trichoderma* come simbionte avirulento della pianta, mentre l'altro riguardante l'interazione fungo-fungo per mettere in evidenza le porzioni genomiche coinvolte nel processo di micoparassitismo. Inoltre, sovrapponendo i dati ottenuti dalle interazioni precedenti si potrebbero determinare i geni chiave dell'azione di biocontrollo svolta dall'isolato oggetto di studio.

- *Analisi interattomica:* potrebbero essere impiegati due isolati fungini differenti ed evolutivamente distanti tra loro, ma aventi lo stesso stile di vita all'interno del solito sistema. In questo modo sarebbe possibile identificare i tratti genomici e le famiglie geniche comuni ad entrambi gli isolati che andrebbero a rappresentare le basi geniche coinvolte in quella precisa/specifica interazione.

Queste analisi potrebbero essere applicate per studiare le basi genetiche ed evolutive degli agenti di biocontrollo appartenenti a questo genere. Inoltre, potrebbero fornire anche strumenti per la selezione massale (sulla base di loci genici) di futuri nuovi agenti di biocontrollo all'interno del genere *Trichoderma*.

6. Bibliografia

Atanasova, L. *et al.* 2013 Comparative transcriptomics reveals different strategies of *Trichoderma* micoparasitism. *BMC Genomics*, 14:121.

Amselem J, Cuomo CA, van Kan JAL, Viaud M, Benito EP, *et al.* 2011. Genomic Analysis of the Necrotrophic Fungal Pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet* 7(8): e1002230.

Bairoch *et al.* 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 1;28(1):45-8.

Baroncelli, R. *et al.* 2014 Draft Genome Sequence of *Colletotrichum acutatum* *Sensu Lato* (*Colletotrichum fioriniae*). *Genome Announcement* Vol.2, No.2.

Bennet, J.W 1997 White pepper: Genomics and Filamentous fungi. *Fungal Genetics and Biology* 21, 3–7.

Bennet, J.W and Arnold, J 2001 Genomics for fungi. *The mycota VIII*, 267-297.

Bernstein ME, Carroll GC (1977). Internal fungi in old-growth Douglas fir foliage. *Can. J. Bot.*, 55: 644-653.

Bissett, J., Szakacs, G., Nolan, C.A., Druzhinina, I., Kullnig-Gradinger, C., Kubicek, C.P., 2003. New species of *Trichoderma* from Asia. *Canadian Journal of Botany* 81, 570–586.

Brown, N.A *et al.* 2012 The Predicted Secretome of the Plant Pathogenic Fungus *Fusarium graminearum*: A Refined Comparative Analysis. *Plos One*, Vol.7, Issue 7.

Carroll, G.C. 1991. Fungal associates of woody plants as insect antagonist in leaves and stems. In P. Barbosa, V.A. Krischik, and C.G. Jones (eds.), *Microbial Mediation of Plant-herbivore Interactions*. Wiley & Sons, New York, pp. 253-271.

Chang YC, Baker R., Kleifeld O., Chet I., 1986. increased growth of plants in the presence of the biological control agent *Trichoderma harzianum*. *Plant Disease* 70:145-148.

Chet I. (1987) *Trichoderma* application, mode of action and potential as biocontrol agent of soilborne plant pathogenic fungi. In: I. Chet (ed.). *Innovative Approaches to Plant Disease Control*. Wiley, New York, pp. 137-160.

Clay. K. 1990. Fungal endophytes of grasses. *Ann. Rev. Ecol. Syst.* 21: 275-295.

- Cuomo,C.A. and Birren,B.W. 2010. Chapter 34 – The Fungal Genome Initiative and Lessons Learned from Genome Sequencing. *Methods in Enzymology*, Vol.470, 2:833–855.
- Daguerre,Y *et al.* 2014 Fungal proteins and genes associated with biocontrol mechanisms of soil-borne pathogens: a review. *Fungal biology reviews*.
- Dean,R.A *et al.* 2005 The genome sequence of the rice blast fungus *Magnaporthe grise*. *Nature*,Vol.43:980-986.
- Degenkolb,T *et al.* 2008 Recent advances and future prospects in peptaibiotics. Hydrophobin and micotoxin research and their importance for chemotaxonomy of *Trichoderma* and *Hypocrea*. *Chemistry & Biodiversity* Vol.5 671-680.
- Dujon B. 2006 Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends in Genetics* Vol.22, No.7.
- Druzhinina,I.S *et al.* 2006. The first 100 *Trichoderma* species characterized by molecular data. *Mycoscience*, Volume 47, Issue 2, pp 55-64.
- Druzhinina,I.S *et al.* 2011 *Trichoderma*: the genomics of opportunistic success. *Nature Reviews*, Vol.9:749-759.
- Druzhinina,I.S *et al.* 2012 Novel traits of *Trichoderma* predicted through the analysis of its secretome. *FEMS Microbiology letters*,337,1-9.
- Elad, Y., Chet. I. & Henis, Y. 1982: Degradation of plant pathogenic fungi by *Trichoderma harzianum*. *Canadian Journal of Microbiology* 28, 719-725.
- Fiorini,L *et al.* 2015 Tomato plant responses network behind the growth promotional effect induced by a novel beneficial isolates of *Trichoderma harzianum* (unpublished).
- Friedl,M.A. and Druzhinina,I.S 2012 Taxon-specific metagenomics of *Trichoderma* reveals a narrow community of opportunistic species that regulate each other's development. *Microbiology*, 158, 69-83.
- Galagan,J.E *et al.* 2005 Genomics of the fungal kingdom: Insights into eukaryotic biology. *Genome Res.* 15, 1620-1631.
- Galagan,J.E *et al.* 2003 The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* Vol.422, 859-868
- Ghisalberti E.L.,Sivasithamparam K. (1991) Antifungal antibiotics produced by *Trichoderma* spp. *Soil Biology & Biochemistry* 23: 1011-1020.

- Goffeau,A *et al*, 1996. Life with 6000 genes. *Science*, Vol.25;274(5287):546, 563-7.
- Grigoriev,I.V 2013 Mycosm portal: gearing up for 1000 fungal genomes, *Nucleic Acids Research*,1–6, doi:10.1093/nar/gkt1183.
- Grigoriev,I.V 2014 A changing landscape of fungal genomics. *The Ecological Genomics of Fungi*, First Edition, Edited by Francis Martin, 3-20.
- Hagan, A., Wallisch, S., Radl, V., Munch, J.C., Schloter, M., 2007. A new cultivation independent approach to detect and monitor common *Trichoderma* species in soils. *Journal of Microbiological Methods* 69, 86–92.
- Harman G.E. (2000). Myths and dogmas of biocontrol. Changes in perceptions derived from research on *Trichoderma harzianum* T22. *Plant Dis.* 84: 377-393.
- Harman G.E., Petzoldt R., Comis A., Chen J. (2004). Interaction between *Trichoderma harzianum* strain T22 and maize inbred line Mo 17 and effect of this interaction on disease caused by *Phitium ultimum* and *Colletotrichum graminicola*. *Phytopathology.* 94:147-153.
- Harman, G.E., Howell, C.R., Viterbo, A., Chet, I. & Lorito, M. 2004: *Trichoderma* species-opportunistic, avirulent plant symbionts. *Nature Review Microbiology* 2: 43-56.
- Harman, G.E. 2006: Overview of mechanisms and uses of *Trichoderma* spp. *Phytopathology* 96, 190-194
- Hermosa,R *et al*. 2012 Plant-beneficial effects of *Trichoderma* and of its genes. *Microbiology*, 158, 17-25
- Holt,C. and Yandell,M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, **12:491**.
- Horton P, Park KJ, Obayashi T, *et al*. 2007 WoLF PSORT: protein localization predictor. *Nucleic Acids Research* 35:585-7
- Howell C.R. (2003) Mechanisms employed by *Trichoderma* species in the biological control of plant disease: the history and evolution of current concepts. *Plant Disease* 87: 4-10.
- Karlsson,M *et al*. 2015 Insights on the evolution of mycoparasitism from the genome of *Clonostachys rosea*. *Genome biology and evolution*, 7(2):465-480
- Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, *et al*. 2011 Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*.

Genome biology, 12:R40.

Laducq,J.B 2014 Ecological genomics of adaptation and speciation in fungi. *Ecological genomics: ecology and the evolution of genes and genomes* 49-72

López-Mondéjar,R *et al.* 2010 Mycoparasitism-related genes expression of *Trichoderma harzianum* isolates to evaluate their efficacy as biological control agent. *Biological control*, doi: 10.1016/j.biocontrol.2010.10.03

Lorito, M., Woo, S.L., Harman, G.E. & Monte, E 2010: Translational Research on *Trichoderma*: From 'Omics to the Field. *Annual Review of Phytopathology* 48, 395-417.

Ma LJ, van der Does HC, Borkovich KA, *et al.* 2010 Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464:367-373

Martinez,D. *et al.* 2004. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat Biotechnol.*22(6):695-700.

Martinez,D. *et al.* 2008. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei*(syn. *Hypocrea jecorina*). *Nature Biotechnology* 26,553-560.

Haunter,S, *et al.* 2011. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.*306-12.

Matarese,F *et al.* 2012 Biocontrol of *Fusarium* head blight: interactions between *Trichoderma* and mycotoxigenic *Fusarium*. *Micorbiology*, 158, 98-106

Miranda MEA, Estrella AH, Cabriales JJP, 2006. Colonization of the rhizosphere, rhizoplane and endorhiza of garlic (*Allium sativum* L.) by strains of *Trichoderma harzianum* and their capacity to control allium white-rot under field conditions. *Soil Biology & Biochemistry* 38:1823-1830

Moncini,L *et al.* 2013 *Trichoderma* su pomodoro: confermata l'azione biostimolante. *L'informatore agrario* 44, 2-5

Morozova,O *et al.* 2008 Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92, 255–264

Mukherjee,M *et al.* 2012 *Trichoderma*-plant-pathogen interactions: advances in genetics of biological control. *Indian J Microbiol* Oct-Dec, 52, 522-529

Mukherjee,P.K *et al.* 2013 *Trichoderma* research in the genome era. *Annu. Rev. Phytopathol.* 51, 105–129

O'Connell RJ, Thon MR, Hacquard S, *et al.* 2012 Lifestyle transitions in plant pathogenic

Colletotrichum fungi deciphered by genome and transcriptome analyses. *Nature genetics* 44(9):1060-1065

Parra, G *et al.* 2007 CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, Vol. 23 no. 9 2007, 1061–1067

Petrini O (1991). Fungal endophytes of tree leaves. In Andrew IA, Hirano SS (eds) *Microbial Ecology of Leaves*, Springer-Verlag, New York, pp 179-197.

Petrini P (1986). Taxonomy of endophytic fungi of aerial plant tissues. In: Fokkema NJ, Van Den Hevel J (eds). *Microbiology of the phytosphere*. Cambridge University Press, Cambridge, pp. 175-187

Quevillon E, Silventoinen V, Pillai S, *et al.* 2005 InterProScan: protein domains identifier. *Nucleic Acids Research* 33:116-120

Rajagopal K. (2012) Diversity of endophytic fungi in some tropical medicinal plants.

Rifai, M.V., 1969. A revision of the fungus *Trichoderma*. CMI Mycological Paper, 116pp.

Rodrigues, K.F., Samuels G. 1999. Fungal endophytes of *Spondias mombin* leaves in Brazil. *Journal of Basic Microbiology*, 39: 131-135.

Rodriguez R.J., White J.F., Arnold A.E., Redman R.S. 2009. Fungal endophytes: diversity and functional roles. *New Phytologist*, 182(2): 314-330.

Rye, C.S and Withers, S.G 2000. Glycosidase mechanisms. *Current Opinion in Chemical Biology*, 4:573–580.

Saikkonen, K., S. H. Faeth, M. Helander, and T. J. Sullivan. 1998. Fungal endophytes: A continuum of interactions with host plants. *Ann. Rev. Ecol. System*, 29:319-343.

Samuels, G.J., 2006. *Trichoderma*: systematics, the sexual state, and ecology. *Phytopathology* 96, 195–206.

Sarrocco, S *et al.* 2004. Are terrestrial Ascomycetes lacking in alginate lytic activity? *J. Gen. Appl. Microbiol.*, 50, 229-234.

Sarrocco, S *et al.* 2013 *Trichoderma harzianum* 6776, a promising biocontrol agent and plant growth promoter. *Biocontrol of plant pathogens in sustainable agriculture*. IOBC/wprs Bulletin Vol.86, 189-194

Sarrocco,S *et al.* 2013 Biological control of Fusarium Head Blight under field conditions. *Biocontrol of plant pathogens in sustainable agriculture*. IOBC/wprs Bulletin Vol.86, 95-100

Schardl,C.L *et al.* 2013 Plant-Symbiotic Fungi as Chemical Engineers: Multi-Genome Analysis of the Clavicipitaceae Reveals Dynamics of Alkaloid Loci. *Plos Genetics*, Vol.9, Issue 2

Shendure,J. and Ji,H 2008. Next-generation DNA sequencing. *Nature Biotechnology*, **26**, 1135 - 1145

Shoresh M., Harman G.E., Mastouri F. (2010) Induced systemic resistance and plant responses to fungal biocontrol agents. *Annual Review of Phytopathology* 48:21-43

Stanke,M *et al.* 2004 AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research*, Vol. 32, Web Server issue W309–W312

Stone J.K., Bacon C.W., White J.E., 2000. An overview of endophytic microbes: endophytism defined. In: *Microbial Endophytes* (Bacon C.W., White J.F., eds.), Marcel Dekker, New York: 1-29.

Studholme, D.J *et al* 2013. Investigating the beneficial traits of *Trichoderma hamatum* GD12 for sustainable agriculture-insights from genomics. *Front Plant Sci*,30;4:258

Suarez,M.B *et al.* 2007 Characterization of genes encoding novel peptidases in the biocontrol fungus *Trichoderma harzianum* CECT 2413 using the TrichoEST functional genomics approach. *Curr Genet* 51:331–342

Ter-Hovhannisyan V, Lomsadze A, Chernoff Y, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Research*.

Verma M., Brar S. K., Tyagi R.D., Surampalli R. Y., Valéro J.R. (2007) Antagonistic fungi, *Trichoderma* spp.: Panoply of biological control. *Review of Biochemical Engineering Journal* 37 (2007) 1–20

Vinale F., Sivasithamparam K.,Ghisalberti E.L., Marra R.,Woo S.L., Lorito M. (2008) *Trichoderma* plant pathogen interactions. *Soil Biology & Biochemistry* 40: 1–10.

Viterbo,A. and Chet,I 2006 *TasHyd*, a new hydrophobin gene from the biocontrol agent *Trichoderma asperellum*, is involved in plant root colonization. *Molecular plant pathology*, 7(4), 249-258

White, T. J., Bruns, T., Lee, S. & Taylor, J. 1990. Amplification and direct sequencing of

fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols : A guide to Methods and Applications* (ed. M. A. Innis, D. H. Gelfand, J. J. Sninsky & T. J. White), pp. 315±322. Academic Press : San Diego, U.S.A

Woo, S.L., Scala, F., Ruocco, M. & Lorito, M. 2006: The molecular biology in the interactions between *Trichoderma* spp., phytopathogenic fungi, and plants. *Phytopathology* 96, 181-185.

Wood,V *et al.* 2002 The genome sequence of *Schizosaccharomyces pombe*. *Nature*,Feb 21;415(6874):871-80.

Xie,B *et al.* 2014 Comparative genomics provide insights into evolution of *Trichoderma* nutrition style. *Genome Biology and Evolution*, doi:10.1093/gbe/evu018.

Xue,M *et al.* 2012 Comparative analysis of the genomes of two field isolates of rice blast fungus *Magnaporthe oryzae*. *Plos genetics* Vol.8, Issue 8, 2-12

Yedidia I., Benhamou N., Chet I. (1999). Induction of defense responses in cucumber plants (*Cucumis sativus* L.) by the biocontrol agent *Trichoderma harzianum*. *Appl. Environ. Microbiol.* 65:1061- 1070.

Yin,Y *et al.* 2012 dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, Vol. 40,W445–W451

Zerbino,D.R. *et al.* 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.*18(5):821-9.

Zhao *et al.* 2013 Correction: Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* doi:10.1186/1471-2164-15-6