UNIVERSITA' DEGLI STUDI DI PARMA

Dottorato di ricerca in Biochimica e Biologia molecolare

Ciclo XXIV

Identificazione funzionale su scala genomica degli attivatori trascrizionali di due funghi filamentosi simbiotici

Coordinatore:
Chiar.mo Prof. Andrea Mozzarelli

Tutor:
Chiar.mo Prof. Angelo Bolchi

Dottoranda: Elisabetta Levati

Sommario

Introduzione **1**
1. I funghi ectomicorrizici 1
1.1 *Tuber melanosporum* 3
1.1.2 Il genoma di *Tuber melanosporum* 7
1.2 *Laccaria bicolor* 9
1.2.1 Il genoma di *Laccaria bicolor* 10
1.3 *Populus trichocarpa* 11
2. I fattori trascrizionali 12
2.1 *Transcriptional activator trap* 21
3. Le proteine nucleari 24
3.1 *Nuclear transportation trap* 26
4. *Moonlighting protein* 27
5. I motivi di legame dei fattori di trascrizione 32
5.1 *Bacterial one-hybrid* 35
Scopo della ricerca **39**
Materiali e metodi **41**
Ceppi di lievito 41
Ceppi di batteri 41
Vettori 42
Terreni per la crescita dei batteri 44
Terreni per la crescita dei lieviti 45
Transcriptional activator trap 48
Nuclear transportation trap 60
Bacterial one-hybrid 62
Risultati e discussioni **70**
1. Analisi *in silico* dei fattori di trascrizione codificati dal genoma di *Tuber melanosporum* 70
2. *Screening* genomico funzionale degli attivatori trascrizionali di *Tuber melanosporum* 75
3. *Nuclear transportation trap* 87
4. Analisi della fosfoadenosina-fosfosolfato reduttasi 91
5. *Bacterial one-hybrid* 97

6. Analisi *in silico* dei fattori di trascrizione codificati dal genoma di *Laccaria bicolor* 106
7. *Screening* genomico funzionale degli attivatori trascrizionali di *Laccaria bicolor* 108

Conclusioni **118**
Bibliografia **120**
Appendice **128**

Introduzione

1. I funghi ectomicorrizici

Esistono numerose specie di funghi che popolano l'intero ecosistema e che differiscono tra loro per lo stile di vita. Ci sono organismi saprotrofi, mutualisti o patogeni. Nello studio condotto ci si è focalizzati sull'analisi del genoma di funghi simbiotici ectomicorrizici.

In generale per simbiosi mediante la formazione della micorriza si intende l'instaurarsi di una connessione tra le ife filamentose del fungo e le radici della pianta. I funghi che appartengono a questa categoria sono molteplici e si ritrovano in habitat che vanno da zone boreali, temperate, montane, foreste tropicali e si stima che circa il 95% delle piante siano colonizzate dai funghi micorrizici. I funghi estendono i miceli sino al contatto con le radici della pianta ospite per formare una fitta rete per lo scambio dei nutrienti tra i due organismi. Dopo il contatto il fungo deve essere in grado di eludere le difese della pianta e deve possedere tutta una serie di proteine che gli consentano di metabolizzare i composti del suolo per produrre i metaboliti di cui necessita la pianta (Martin et al., 2007).

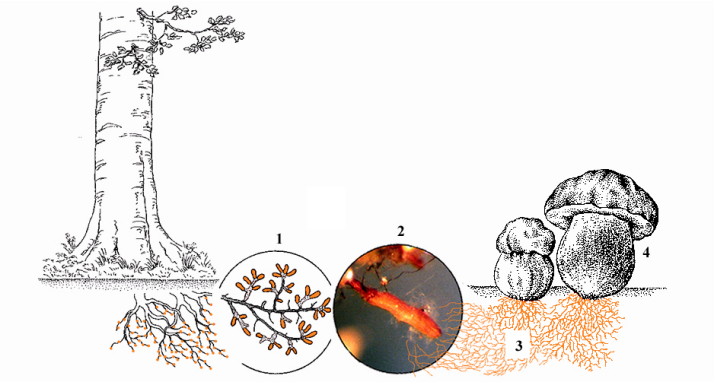


Fig 1.1: Simbiosi tra fungo e pianta (Montecchio, 2006)

I funghi ectomicorrizici, instaurando una simbiosi con le radici di molte piante e creando una fitta rete di scambio di nutrienti, determinano un aumento del benessere e della produttività forestale. Questi funghi sono fondamentali per l'ecosistema terrestre poiché molte piante dipendono da loro per quanto riguarda l'*up-take* di alcuni nutrienti e perchè spesso la simbiosi porta ad un a un beneficio per la pianta anche per quanto riguarda l'aumento della resistenza a patogeni e a stress come elevate temperature o siccità (Kernaghan, 2005).

All'interno dei basidiomiceti e degli ascomiceti ritroviamo i funghi ectomicorrizici che sono in grado di instaurare simbiosi con circa il 3% delle piante da seme.
Il successo nell'instaurazione della simbiosi è dovuto ai benefici tratti dalla pianta nel ricevere dal fungo nutrienti che da sola non è in grado di metabolizzare. I funghi micorrizici hanno un ruolo fondamentale nell'ambiente per quanto riguarda il metabolismo di composti a base di azoto, carbonio e fosforo. La pianta dipende dal fungo per quanto riguarda l'*up-take* dei fosfati, dei composti azotati organici come amminoacidi solubili o derivanti da proteine ma anche di fonti inorganiche di azoto come ammonio e nitrati. Il fungo possiede gli enzimi che gli permettono di convertire queste molecole in metaboliti assimilabili dalla pianta. Attraverso la formazione dell'ectomicorriza, struttura specializzata per la simbiosi, il fungo riesce, mediante appropriati trasportatori, a fornire alle radici i nutrienti di cui necessita la pianta. La pianta a sua volta, instaurando una interazione mutualistica, fornisce al fungo zuccheri che derivano dalla fotosintesi. Da studi condotti su questi organismi è emerso che lo stimolo per l'instaurazione della simbiosi parta dai funghi poiché la pianta è in grado di sopravvivere anche senza la simbiosi mentre spesso questi funghi dipendono completamente della pianta per quanto riguarda le fonti di carbonio (Plett and Martin, 2011).
Durante la simbiosi i filamenti ramificati del fungo si estendo nel terreno e il loro sviluppo è indirizzato all'incontro con la radice della pianta; una volta contattata ramificano tra le cellule dello strato esterno della radice e formano una fitta rete di ife. Le ife riescono a penetrare la cuffia radicale e l'invasione continua sino al raggiungimento delle cellule epidermiche. A seguito del contatto tra le cellule corticali e le ife si hanno delle modificazioni morfologiche che portano alla formazione di una reticolo detto reticolo di Harting in cui le ife raggiungono lo spazio apoplastico formando la struttura responsabile dello scambio di nutrienti tra i due organismi. Si stima che circa il 20% delle fonti di carbonio assimilate dalla pianta possa essere ceduto al fungo che metabolizza glucosio e saccarosio in modo da creare un gradiente che favorisca un continuo *up-take* di fonti di carbonio derivanti dalla pianta. Il fungo invece cede alla pianta circa il 70% dell'azoto e fosforo assorbito. La simbiosi garantisce una ottima crescita della pianta anche in terreni poveri di nutrienti.
Non è noto se i funghi simbiotici possiedono un insieme di geni responsabili della simbiosi o se ognuno ha sviluppato meccanismi diversi per il mutualismo. Risposte interessanti potranno esserci fornite dallo studio dei genomi dei due funghi simbiotici *Tuber melanosporum* e *Laccaria bicolor* unitamente allo studio del genoma di una delle

piante con cui possono instaurare la simbiosi che è il pioppo *Populus trichocarpa.* Si cercherà di far emergere quali fattori guidano questo meccanismo di simbiosi, quali sono le proteine che promuovono la simbiosi con particolare interesse per quanto riguarda le proteine secrete che sono responsabili della comunicazione con la pianta e i fattori di trascrizione che mediano la regolazione dell'espressione genica (Plett and Martin, 2011).

1.1 Tuber melanosporum

Tuber melanosporum, il tartufo nero pregiato del Perigord, é un ascomicete filamentoso simbiotico ipogeo. Domina, insieme al tartufo bianco di Alba *Tuber magantum*, il mercato dei tartufi poiché sono considerati una vera e propria prelibatezza culinaria grazie al loro specifico sapore e inconfondibile aroma. Il valore commerciale del *Perigord* è compreso tra i 300 e 1000 €/kg.

T. melanosporum si sviluppa in terreni calcarei nel sud Europa e in particolare in Francia, Italia e Spagna. È presente in climi temperati, mediterranei e continentali e sono esclusi i climi tropicali, secchi (con media annuale di precipitazioni inferiori ai 350mm) e climi molto caldi. I terreni in cui si sviluppa sono brecciosi, ben drenati, ricchi di carbonati di calcio, ferro e rame, mentre sono poveri di sostanza organica, azoto, fosforo e potassio, il pH del terreno è alcalino. Le tartufaie di solito si sviluppano ad altitudini comprese tra i 100 e i 1000 metri su altipiani, zone collinari o boschi non fitti. Le tartufaie di *T. melanosporum* sono ben riconoscibili per la formazione nel terreno sottostante la pianta di una zona detta bruciata, cava o pianello priva di vegetazione poiché il tartufo produce un prodotto tossico che impedisce la germinazione delle piante erbacee. Le piante con cui instaura la simbiosi sono soprattutto querce, noccioli o carpini neri. Il tartufo viene raccolto da dicembre a marzo.

La distribuzione naturale e genetica delle popolazioni di questo tartufo nero è stata condizionata maggiormente da 5 fattori:

- Distribuzione delle piante ospiti
- Spore disperse da animali
- Fattori ecologici limitanti (es: terreni calcarei e climi temperati)
- Barriere geografiche (es: il mar Mediterraneo che ne limita l'espanzione verso il Nord Africa)
- Eventi storici

Analisi filogeografiche delle popolazioni di *Tuber* hanno evidenziato che il fattore principale che ha influenzato la struttura della popolazione di questo tartufo è stata l'espansione post-glaciale delle piante ospiti (Martin et al., 2010).

I funghi simbiotici colonizzano le piante mediante la formazione di una struttura specializzata, l'ectomicorriza, nella quale avviene lo scambio di nutrienti tra il fungo e l'apparato radicale: essi aumentano l'assunzione di azoto e altri nutrienti dal terreno e li trasferiscono alle piante in cambio fonti di carbonio derivanti dai prodotti della fotosintesi. A causa della loro capacità simbiotica si pensa che i funghi micorrizici giochino un importante ruolo nell'evoluzione e nel mantenimento della maggior parte delle attuali piante arboree. Oltre alla fase simbiotica, passaggio obbligato perché avvenga la formazione del corpo fruttifero, il loro ciclo vitale presenta una fase vegetativa nella quale si ha una crescita indipendente e avviene la formazione del micelio, una struttura caratterizzata da filamenti (ife) ramificati e intrecciati, e una fase riproduttiva (corpo fruttifero) nella quale avviene la produzione di spore in una particolare ifa cava detta asco da cui escono o vengono espulse a maturità con meccanismi diversi.

La fruttificazione dell'ectomicorriza di *Tuber spp*. dipende da un notevole numero di variabili, tra cui metaboliti e segnali prodotti dalla pianta ospite, livello di nutrienti nel substrato e condizioni ambientali (umidità e temperatura). Tutte queste variabili modificano l'espressione genica determinando lo *shift* morfogenetico che porta al corpo fruttifero finale, anche detto tartufo di elevata tipicità e valore commerciale.

Le spore che germinano in primavera producono un micelio vegetativo aploide che cresce e si estende nel terreno e nella rizosfera sino a colonizzare le radici delle piante ospiti e instaurare la simbiosi mutualistica. Le ife che crescono vicine alla micorriza possono aggregarsi per iniziare la formazione di un prematuro corpo fruttifero. L'ultima fase dello sviluppo del corpo fruttifero, struttura multicellulare molto complessa, avviene durante l'inverno e completa il ciclo vitale. Mentre le ascospore vengono prodotte dalle ife dicariotiche a seguito della meiosi i tessuti formanti il corpo fruttifero vengono prodotti da ife aploidi. Lo sviluppo del corpo fruttifero richiede particolari condizioni ambientali e inoltre il micelio deve aver raggiunto un certo grado di competenza prima di potersi differenziare. I fattori più importanti sono: nutrienti, luce, temperatura, areazione e pH. I diversi tessuti che formano il corpo fruttifero sono il risultato di un processo di differenziazione che porta alla formazione degli aschi che racchiudono le meiospore. Le ascospore dopo meiosi sono disperse tramite animali, tra cui micetofagi, che possono rompere gli aschi liberando le ascospore oppure

possono ingerirle e disperderle con le feci. A differenza dei funghi epigei le cui spore possono essere propagate sulle lunghe distanze dal vento, le spore di *Tuber,* fungo ipogeo, vengono disperse solo su corte distanze (Martin et al., 2010).

Le dimensioni del tartufo sono variabili e generalmente non superano i 200-300 grammi ad eccezione di casi particolari. Il corpo fruttifero è tondeggiante ma in alcuni casi può avere una forma irregolare e lobata, il peridio è nero e può presentare delle striature rossastre mentre la gleba da chiara diventa bruna con venature biancastre nei tartufi maturi.

I meccanismi molecolari che stanno alla base della formazione del corpo fruttifero sono poco noti. Alcuni studi hanno cercato di mettere in luce quali fossero i passaggi principali che determinano la transizione dalle ife vegetative alle strutture multicellulari rappresentate dai corpi fruttiferi in ascomiceti filamentosi modello. La maggior parte degli studi sono stati condotti in organismi quali *Aspergillus nidulans* e *Sordaria macrospora*, funghi omotallici capaci di formare corpi fruttiferi e facilmente trattabili con approcci genetici e con cui è possibile eseguire l'analisi di mutanti. Lo sviluppo delle ife vegetative è spesso sufficiente per la sopravvivenza, ma il micelio può anche costituire il punto di partenza per il differenziamento in diversi tipi cellulari in risposta al microambiente che circonda il fungo (Ugalde, 2006). I corpi fruttiferi si generano in presenza di diversi stimoli ambientali e di segnali endogeni autoregolatori. I fattori esterni generalmente includono una serie di parametri tra i quali la disponibilità di nutrienti, la luce, la pressione dell'aria, il contatto con la superficie, l'osmolarità e il pH. Tra i fattori endogeni sono presenti i componenti del metabolismo primario, feromoni o sostanze ormone-simili e lo stato redox della cellula. La percezione di questi segnali determina il coinvolgimento di diversi geni che codificano per proteine G, recettori e fattori di trascrizione. Ad esempio è noto che il meccanismo di percezione della luce di *A. nidulans* include il recettore citocromo della luce rossa (Blumenstein et al., 2005). I più importanti sistemi di percezione del segnale individuati sono i recettori per le proteine che legano la guanina (GPCRs) e le corrispondenti proteine G. Dei 16 GPCRs di *A. nidulans*, GprA, GprB e GprD, sono coinvolti nella regolazione della fase riproduttiva (Yu, 2006). Le due principali vie di traduzione del segnale, la cascata MAPK e la cascata cAMP-PKA, funzionano in modo coordinato per regolare il processo di differenziazione cellulare mediante l'attivazione o la repressione di numerosi fattori trascrizionali appartenenti a diverse classi. Questi regolano a loro volta la trascrizione di geni che codificano per enzimi coinvolti nella biogenesi della parete cellulare e nel metabolismo e la trascrizione di geni richiesti per l'organizzazione e la

struttura del citoscheletro. L'espressione di alcuni di questi geni comporta anche una regolazione ad un livello post-trascrizionale. CpeA di *A. nidulans* (Scherer et al., 2002) e APP di *S. macrospora* (Nowrousian et al., 2007), che codificano rispettivamente per una catalasi/perossidasi e per una proteina del peritecio, sono entrambi fortemente espressi nella fase di crescita precoce ma l'accumulo delle corrispondenti proteine è lento. Questo potrebbe riflettere un controllo a livello traduzionale della sintesi o post-traduzionale della stabilità delle proteine. La proteina COP9 che regola la degradazione delle proteine ubiquitinina-dipendente e gli adattatori per la ligasi dell'ubiquitinina E3 come GrrA, sono importanti per la formazione del corpo fruttifero e per l'ascosporogenesi (Busch et al., 2004) (Krappmann et al., 2006). Ciò suggerisce che il processo potrebbe richiedere uno specifico insieme di proteine che è assente nelle cellule vegetative. È evidente che lo sviluppo di questa struttura multicellulare richiede una specifica integrazione di diversi processi biologici fondamentali e i fattori di trascrizione sono i componenti meno conosciuti di quest'ultimi. La via riproduttiva negli ascomiceti costituisce un valido sistema sperimentale per lo studio dei meccanismi e della regolazione dei processi di sviluppo che sono normalmente più complessi in animali e piante. In particolare, *T. melanosporum* rappresenta un buon candidato per l'analisi dello stile di vita simbiotico e per la comprensione dei meccanismi di formazione di un corpo fruttifero che cresce sottoterra (ipogeo).

L'analisi dei genomi dei funghi diventa il punto di partenza per giungere ad ulteriori conoscenze nei processi di crescita dei funghi filamentosi e di altri eucarioti multicellulari. Negli ultimi 10 anni sono state determinate le sequenze genomiche di oltre 60 ascomiceti e ciò ha permesso di individuare nuove classi di geni e di delineare alcuni dei determinanti molecolari che differenziano un fungo saprotrofo da un fungo patogeno. Sono stati isolati e caratterizzati gli omologhi di *Tuber* di diversi geni coinvolti in processi quali la sintesi della parete, il metabolismo dei carboidrati e dell'azoto. Malgrado l'innegabile interesse dei tartufi come "biofertilizzatori" e la crescente visibilità scientifica da essi acquisita, rimangono inesplorati dal punto di vista genomico sia lo stile di vita simbiotico/ipogeo, sia i processi responsabili della formazione del corpo fruttifero. A questo proposito due classi di proteine che attirano maggior interesse sono le proteine secrete e di parete e i fattori trascrizionali. Sono particolarmente interessanti poiché l'interazione tra l'ifa di un fungo e l'ambiente esterno, sia esso il suolo, un (micro)organismo omologo o competitore, o la radice di una pianta, è mediata dalle proteine che il fungo secerne o espone sulla sua superficie e dai fattori di trascrizione poiché queste interazioni modificano i programmi di espressione genica per la

differenziazione del fungo. Oltre al loro indiscutibile valore come "modelli" per la decifrazione dell'informazione genetica di organismi più complessi (tra cui l'uomo), i genomi dei funghi rappresentano le piattaforme conoscitive per la comprensione del ruolo da essi svolto nel controllare gli equilibri della biosfera e nella conversione della biomassa (e.g., lignina e cellulosa), per lo sviluppo di nuove strategie di salvaguardia ambientale e per lo sfruttamento biotecnologico delle risorse biologiche in essi contenute. Particolare interesse in ambito agro-ambientale (e alimentare) è rivestito appunto dai funghi simbiotici che grazie al loro ruolo di "bio-fertilizzatori" offrono straordinarie opportunità per una agricoltura sostenibile e a basso impatto chimico.
La crescente domanda per questo tartufo e il forte impatto che questo organismo ha sia a livello economico che ecologico ha portato ad un intenso sforzo per cercare di comprendere i meccanismi molecolari che stanno alla base dello sviluppo del corpo fruttifero e dell'instaurazione della simbiosi con la pianta.

1.1.2 Il genoma di *Tuber melanosporum*

Il genoma di *Tuber*, composto da 125 megabasi, è stato pubblicato nel 2010 e rappresenta il più grande genoma fungino sequenziato sino ad ora. Al suo interno sono state identificate 7500 sequenze codificanti per proteine. Il numero di geni codificanti per proteine, nonostante la dimensione del genoma, è paragonabile ad altri ascomiceti. La dimensione del genoma non è da implicare ad un evento di duplicazione ma è da attribuire alla presenza di sequenze trasponibili che ricoprono il 58% dell'intero genoma. I trasposoni non sono diffusi uniformemente nel genoma ma sono raggruppati in regioni povere di sequenze geniche (Martin et al., 2010).
Delle 7500 proteine predette 3970, 5596, 5644 mostrano similarità di sequenze significative con proteine rispettivamente di *Saccharomyces cerevisiae*, *Neurospora crassa*, *Aspergillus niger* tuttavia tra *Tuber* e questi organismi si riscontra un basso livello di sintenia che porta alla conclusione che questo fungo abbia una diversa strutturazione del genoma rispetto agli altri ascomiceti sequenziati (Martin et al., 2010).
È stata valutata anche l'espressione dei geni nei diversi stadi del ciclo vitale e da queste analisi emerge che solo il 7,6% dei trascritti è differenzialmente espresso nella micorizza o nel corpo fruttifero rispetto al micelio e che solo 61 trascritti sono unici di uno dei tre stadi del ciclo vitale. Nel tessuto simbiotico risultano sovraespresse proteine coinvolte nell'adesione cellulare, nella detossificazione di metaboliti che la pianta produce come meccanismo di difesa, negli scambi di nutrienti, nella colonizzazione degli apopalsti della radice e proteine che agiscono sulla parete cellulare della pianta. Un processo fondamentale per il successo dell'instaurazione della simbiosi è lo

scambio di nutrienti tra la pianta ospite e il fungo e se paragonato con altri organismi simbiotici o con funghi patogeni il numero dei trasportatori presenti in *Tuber* è minore ma è bilanciato dal fatto che 64 dei 381 trasportatori risultano sovraespressi nella micorriza suggerendo così un aumento del flusso dei carboidrati, oligopeptidi, aminoacidi, e poliammine (Martin et al., 2010).

Una ulteriore caratteristica del genoma di questo organismo è la quasi assenza di copie di geni altamente simili, dei 7500 geni predetti solo sette copie mostrano una similarità maggiore del 90% mentre 30 copie mostrano l'80% di similarità. Anche le famiglie multigeniche sono limitate e comprendono solo il 19% del proteoma, inoltre molte di queste famiglie sono rappresentate da solo due membri. Se paragonate ad altri ascomiceti le famiglie geniche che codificano per trasportatori di metaboliti e enzimi coinvolti nel metabolismo secondario sono più ristrette. Le differenze nell'espansione delle famiglie geniche sono probabilmente la causa dello svilupparsi di differenze nel meccanismo di simbiosi e nella specificità dell'ospite tra *T.melanosporum* e *L. bicolor.* Il genoma compatto di questo fungo può essere una conseguenza della sua specializzazione a differenza di *Laccaria* che sfrutta diversi substrati forniti da potenziali ospiti e che è in grado di svilupparsi in diversi terreni (Martin et al., 2010).

Dato che questo è il primo genoma sequenziato di un fungo produttore di corpi fruttiferi ipogei con un caratteristico sapore sono stati annotati anche enzimi specifici per l'assimilazione dello zolfo inclusa la cistationina liasi e vari enzimi per la degradazione degli amminoacidi che sono responsabili della formazione dei composti volatili che conferiscono al tartufo il suo peculiare aroma. Nonostante sia un fungo ipogeo sono stati riscontrati anche geni coinvolti nella percezione della luce che si pensa siano responsabili dello sviluppo del corpo fruttifero o della riproduzione sessuale in base alle condizioni stagionali. L'analisi del genoma ha verificato il potenziale alimentare di *Tuber* poiché rappresenta una specie fungina non allergenica in quanto possiede solo quattro geni con una similarità maggiore dell'80% rispetto a allergeni noti e poiché non produce nessuna micotossina (Martin et al., 2010).

1.2 Laccaria bicolor

Laccaria bicolor è un fungo micorrizico basidiomicete che cresce nelle zone temperate del globo, nel Nord America e nelle foreste eurasiatiche, in tarda estate ed

autunno. È caratterizzato da un cappello convesso di diametro variabile (4-5 cm), colorato con varie sfumature tra il marrone e l'arancio, con superficie squamuloso-furfuracea. Il gambo si presenta cilindrico e slanciato, dello stesso colore del cappello e con micelio violetto alla base. Comunemente a tutti i funghi simbiotici, crea un rapporto mutualistico con la pianta mediante formazione di una struttura specializzata, l'ectomicorriza, nella quale avviene lo scambio di nutrienti tra il fungo e l'apparato radicale*:* il fungo sfrutta gli zuccheri provenienti dalla fotosintesi mentre la pianta ospite sfrutta essenzialmente i minerali, l'azoto organico ed aumenta l'*uptake* dell'acqua*.*

Laccaria bicolor è stato il primo fungo simbiotico interamente sequenziato*.* Prima del sequenziamento del genoma, rappresentava la specie fungina favorita per gli studi sull'ectomicorriza. Al contrario di altri funghi simbiotici infatti, si può far crescere su substrati sintetici (mediante basidiospore o campioni di tessuto) o insieme alle radici di alberi con cui effettua l'interazione mutualistica (pioppi ed altre latifoglie), permettendo studi sulla sua fisiologia, biochimica ed interazione con la pianta ospite in condizioni controllate.

Laccaria produce meiospore bianche e spinose (echinulate) che germinano nelle vicinanze delle radici della pianta ospite producendo micelio aploide monocariotico. I due complessi dei fattori del *mating-type* controllano la compatibilità sessuale e regolano il mantenimento dello stato dicariotico. La fusione di miceli monocariotici aploidi sessualmente compatibili porta alla formazione del micelio dicariotico (diploide). Le ife diploidi sono predominanti nelle strutture vegetative di *Laccaria* e di molti altri basidiomiceti. In determinate condizioni ambientali, il micelio filamentoso diploide attaccato alle radici della pianta ospite produce il corpo fruttifero senza che sia avvenuta meiosi. Il micelio sia aploide che diploide è capace di crescita indefinita nel suolo, assicurando il mantenimento e la duplicazione del genotipo di ogni ploidia. La simbiosi è generalmente indotta dai miceli diploidi. Alterazioni drammatiche dell'espressione dei geni sia di pianta che di fungo sono richieste durante la formazione dell'ectomicorriza. Anche se tutt'oggi esse sono ancora sconosciute, è certo che sia fondamentale lo scambio di metaboliti nella rizosfera, il bilancio ormonale e meccanismi di segnalazione che coinvolgono il Ca^{2+} che sono regolati da chinasi. Durante la fase simbiotica risultano sovraespressi geni che codificano per proteine secrete che possono agire andando a bloccare l'azione delle proteasi della pianta o possono modificare l'interfaccia tra fungo e pianta. È inoltre indispensabile l'inibizione dei meccanismi di difesa della pianta per garantire il successo della simbiosi.

Oltre alla fase simbiotica, il ciclo vitale del fungo prevede altre due fasi: la fase vegetativa nella quale si ha crescita indipendente con produzione di miceli (ife), e la fase riproduttiva (corpo fruttifero) nella quale avviene la produzione di spore, in una particolare ifa cava detta asco da cui escono o vengono espulse a maturità con meccanismi diversi. Le ife sono strutture filamentose ramificate e intrecciate, si sviluppano vicino alle radici della pianta ospite per poi aggregarsi e formare il corpo fruttifero iniziale (Martin et al., 2008).

1.2.1 Il genoma di *Laccaria bicolor*

La sequenza completa di 65 Mb è stata pubblicata nel 2008 (Martin et al., 2008). Il genoma contiene circa 20614 geni che codificano per proteine, un numero elevato di trasposoni e sequenze ripetute. Nel genoma non ci sono evidenze di duplicazioni su larga scala però si possono osservare duplicazioni in tandem tra famiglie multigeniche. Anche in *Laccaria* si ritrova un elevato numero di sequenze trasponibili che ricoprono il 21% del genoma.

Delle circa 20000 proteine predette il 70% mostra similarità di sequenza con proteine note, la maggior parte delle quali si ritrovano nei genomi sequenziati di altri basidiomiceti. Si nota inoltre una marcata espansione delle famiglie geniche che codificano per proteine che hanno un ruolo nell'interazione proteina-proteina e nella trasduzione del segnale inoltre molti di questi trascritti risultano sovraespressi nella micorriza e nel corpo fruttifero suggerendo un ruolo nella differenziazione dei tessuti.

Sono state ritrovate inaspettate peculiarità genomiche di questo fungo tra cui una batteria di proteine secrete (SSPs) molto interessanti poiché l'82% sono uniche di *Laccaria* mentre il 67% hanno funzione sconosciuta. I profili trascrittomici inoltre evidenziano come molte di queste proteine siano espresse preferenzialmente nell'ectomicorriza sottolineando un ruolo attivo di queste proteine per la comunicazione con la pianta ospite e per l'instaurazione della simbiosi.

Anche se il genoma contiene numerosi geni codificanti per enzimi idrolitici si osserva una significativa riduzione di enzimi coinvolti nella degradazione degli oligosaccaridi presenti nella parete cellulare della pianta, questo è indice di un adattamento del fungo ad una condizione di simbiosi che ha portato alla perdita di questi geni. I geni con tale funzione presenti nel genoma probabilmente hanno un ruolo di rimodellamento della parete cellulare durante la differenziazione dei tessuti perché risultano maggiormente espressi nel corpo fruttifero o nell'ectomicorriza oppure un ruolo durante la

penetrazione dello spazio apoplastico della radice poiché sovraespressi solo nella micorriza.
Possiede inoltre enzimi capaci di degradare oligosaccaridi diversi da quelli presenti nella parete della pianta che gli permettono la sopravvivenza in modo autonomo nel suolo prima della associazione simbiotica.
È importante sottolineare che *L. bicolor* a differenza di *T. melanosporum* non possiede nel suo genoma il gene che codifica per una invertasi. Quindi questo fungo è completamente dipendente dalla pianta per l'assunzione del glucosio.
L'identificazione dei fattori primari che regolano lo sviluppo simbiotico e le attività metaboliche permette la comprensione del ruolo di ectomicorriza nello sviluppo della pianta e nella sua fisiologia, aprendo le porte alla comprensione del significato biologico di questa simbiosi (Martin et al., 2008) (Martin and Selosse, 2008).

1.3 Populus trichocarpa

Le foreste ricoprono il 30% della superficie terrestre determinando una sostanziale biodiversità e una risorsa per l'umanità con benefici per quanto riguarda la qualità dell'aria e dell'acqua ma anche rappresentando una fonte di legname e carburanti. La crescita e lo sviluppo delle piante forestali coinvolge diversi fattori: crescita secondaria, distribuzione dell'acqua e dei nutrienti su grandi distanze, l'immagazzinamento e la ridistribuzione dei metaboliti durante l'alternarsi delle stagioni e i cambiamenti climatici. Quello che diversifica le piante da altri organismi sono le grandi dimensioni e il lungo tempo di generazione.
Tra le varie latifoglie che sono interessate dalla simbiosi con *Laccaria bicolor,* vi è anche il pioppo *Populus trichocarpa*, del quale è disponibile dal 2006 la sequenza genomica (Tuskan et al., 2006). Questa varietà di pioppo, detta *Black Cottonwood*, cresce in America, principalmente nel Montana, e nell'Europa occidentale, arrivando ad altezze che superano i 100 metri e vivendo per più di 200 anni. Il genere *Populus* è ormai accettato come un modello per la ricerca di base sulle piante arboree per la dimensione ridotta del suo genoma, per la loro rapida crescita e per la facilità di manipolazione. Il sequenziamento tramite shotgun è stato effettuato utilizzando un singolo genotipo eterozigotico femminile detto "Nisqually 1" e con esso si è definito un genoma di circa 485Mb complessive di cui 410Mb sono state assemblate mentre le restanti 75 Mb sono compatibili con l'ipotesi che il 30% del genoma sia rappresentato da eterocromatina. Sono stati assemblati anche i genomi degli organelli, quello dei cloroplasti è di 157 kb e quello mitocondriale di 803kb.

Nel genoma nucleare sono stati identificati 45555 geni codificanti per proteine, 101 in quello dei cloroplasti e 52 in quello mitocondriale. Delle proteine identificate l'89% ha similarità di sequenza con proteine presenti nel database dell'NCBI e il 12% non mostra similarità con proteine di *Arabidopsis*. Il 53% delle proteine mostra una espressione differenziale nei diversi tessuti.
All'interno del genoma ritroviamo l'espansione di alcune famiglie geniche che hanno contribuito all'evoluzione e allo sviluppo di questa specie come ad esempio i geni coinvolti nella formazione della parete di lignocellulosa, i geni per la produzione di metaboliti secondari, geni per la resistenza a patogeni, trasportatori di membrana e geni per la produzione di fitormoni (Tuskan et al., 2006).
Con la sequenza genomica di pioppo si sono potute incrementare le informazioni ottenute analizzando il genoma di *Arabidopsis* riguardo lo sviluppo della corteccia, della radice, della corona ma anche dei nutrienti utilizzati dalla pianta, delle malattie e delle resistenze a vari patogeni. Inoltre grazie al sequenziamento di questa varietà di pioppo, si possono effettuare studi genomici per aiutare la comprensione dei meccanismi di simbiosi tra pianta e fungo.

2. I fattori trascrizionali

I fattori trascrizionali (TF) sono proteine che controllano l'espressione genica e riconoscono specifiche sequenze di DNA chiamate elementi di regolazione in *cis* come promotori, enhancer o altre regioni regolative presenti nei genomi degli eucarioti (Stegmaier et al., 2004). I regolatori trascrizionali eucariotici che attivano la trascrizione sono normalmente composti da un dominio di legame al DNA (DBD) e da un dominio di attivazione (AD). Il DBD garantisce il riconoscimento di queste proteine di specifici siti di legame del DNA e l'AD media l'inizio della trascrizione. I fattori di trascrizione hanno diversi meccanismi di azione e vengono anche influenzati da altri fattori, ad esempio l'interazione con altre proteine come altri fattori di trascrizione, proteine che modificano la cromatina o chinasi. La crescente disponibilità di sequenze genomiche ha permesso uno studio approfondito del repertorio di fattori trascrizionali e ha messo in evidenza la complessità e la diversità di questi meccanismi di regolazione nei diversi organismi.

I fattori di trascrizione vengono classificati in base alla tipologia di dominio di legame al DNA che posseggono ma bisogna considerare che anche altre regioni della proteina possono contribuire alla loro funzione ad esempio possono contenere regioni indispensabili per la dimerizzazione con altre proteine. I TF possono omodimerizzare cioè formare un dimero tra proteine identiche oppure eterodimerizzare cioè formare un dimero tra proteine diverse e il legame al DNA può avvenire prima o dopo l'interazione delle due proteine. L'evento di regolazione diventa quindi dipendente dalla scelta del partner. Il vantaggio derivato dalla dimerizzazione è quello di poter regolare molte funzioni andando a modulare l'interazione tra proteine, questo implica che anche un ridotto numero di fattori di trascrizione può svolgere molti ruoli semplicemente cambiando il suo partner di interazione. La capacità di formare dimeri conferisce molta flessibilità e complessità al meccanismo di regolazione genica. All'interno della sequenza proteica il dominio responsabile della dimerizzazione è generalmente collocato a valle del dominio di legame al DNA ma non è molto conservato. Le maggiori implicazioni di questo fenomeno sono: la possibilità di ottenere espressione differenziale regolata dalla eterodimerizzazione, facilitare l'orientamento e la prossimità dei fattori di trascrizione alla sequenza regolatrice e aumentare la specificità di legame. Dato che un monomero può interagire con diversi TF e formare dimeri con diverse funzioni diventa importante la regolazione della formazione del dimero che può essere influenzata da fattori come la concentrazione del monomero o le modificazioni post-traduzionali dei monomeri. Gli eterodimeri inoltre andando a combinare diversi domini di legame conferiscono al complesso proteico distinte specificità di legame (Amoutzias et al., 2007).

I fattori di trascrizione sono in grado di modulare l'espressione genica sia come attivatori che come repressori in modo dipendente al contesto cellulare in cui si trovano e dai cofattori con cui interagiscono. I fattori di trascrizione non solo contattano e reclutano la RNA polimerasi ma si possono legare a proteine responsabili del rimodellamento della cromatina che influenzano lo stato di modificazione degli istoni e che quindi possono regolare l'accessibilità al DNA e influenzare l'attivazione o repressione dell'espressione genica.

I diversi genomi eucariotici sino ad ora sequenziati mostrano delle differenze nel repertorio dei fattori di trascrizione sia per quanto riguarda la presenza che la prevalenza di queste proteine. Si stima che i fattori di trascrizione occupino una porzione del genoma che varia dallo 0,5 sino al 8% delle sequenze geniche codificanti per proteine. Il numero di TF codificati da un organismo è solitamente proporzionato

alla dimensione e alla complessità del genoma. I fattori di trascrizione eucariotici riconoscono corte sequenze di DNA degenerate e si presuppone che il punto chiave della regolazione sia l'interazione e la cooperatività tra i TF piuttosto che una perfetta specificità di sequenza. I domini di legame presenti nei fattori di trascrizione prevalentemente legano il DNA contattando il solco maggiore ma non mancano esempi di legami anche con il solco minore o di interazioni con i gruppi fosfato.

I fattori di trascrizione sono stati classificati in 5 superfamiglie in base al loro dominio di legame al DNA (Stegmaier et al., 2004):

0- Fattori che contengono altre caratteristiche
1- Fattori contenenti domini basici
2- Domini zinc finger
3- Domini elica-giro-elica
4- Domini β-scaffold

La prima comprende i fattori contenenti domini basici. Questi fattori hanno un DBD caratterizzato da un largo eccesso di residui carichi positivamente che consentono alla proteina di assumere una struttura ad α-elica solamente quando interagisce con il DNA. Sono strettamente correlati ad un dominio di dimerizzazione, una cerniera a leucina (bZIP) oppure ad un motivo elica-loop-elica (bHLH). La dimerizzazione di questi fattori è un requisito fondamentale per il loro legame al DNA e contribuisce in gran parte alla loro specificità di legame (Stegmaier et al., 2004). Alla prima classe appartenente a tale superfamiglia appartengono i fattori di trascrizione che possiedono un motivo a cerniera di leucine (bZIP). Questo motivo è costituito da una α-elica bipartita lunga circa 60-80 amminoacidi che contiene un dominio di legame al DNA all' N-terminale formato da due cluster basici e un dominio di dimerizzazione formato da una elica anfipatica al C-terminale che contiene un residuo di leucina ogni 7 amminoacidi (eptade). Questi fattori possono sia omodimerizzare che etrodimerizzare conferendo così ad ogni fattore di trascrizione la capacità di svolgere funzioni multiple (Borkovich et al., 2004).

Il fattore di trascrizione dimerizza formando due α-eliche parallele superavvolte che consistono di 4 o 5 eptadi. Ogni eptade è formata da sette residui che vengono chiamati a, b, c, d, e, f, g. Gli amminoacidi in posizione a-d-e-g sono responsabili della oligomerizzazione, della stabilità del dimero e della specificità della formazione del dimero. In posizione d troviamo la leucina. Gli amminoacidi presenti in a-d interagiscono con le medesime posizioni presenti sull'elica della proteina con cui dimerizzano e formano un'interfaccia idrofobica importante per la stabilità mentre i

residui b-c-f sono nella regione idrofilica della proteina. I residui chiave per la dimerizzazione non sono idrofobici ma spesso in posizione “a” ritroviamo una asparagina che favorisce la omodimerizzazione o una lisina che favorisce la eterodimerizzazione. I residui in g ed e fiancheggiano la regione idrofobica e sono caratterizzati da residui carichi, tra questi residui e i corrispettivi presenti nell’altra elica si creano interazioni elettrostatiche che possono essere sia repulsive che attrattive e determinano anch’essi l’omodimerizzazione o l’eterodimerizzazione. (Deppmann et al., 2006) (Vinson et al., 2002).

Domini a cerniera a leucina corti conferiscono meno flessibilità alla proteina in quanto la sequenza amminoacidica deve essere ottimizzata in funzione della stabilità del dimero mentre domini lunghi permettono una migliore regolazione poiché possono contenere amminoacidi ottimali sia per la stabilità che per la specificità di interazione con altre proteine (Vinson et al., 2002).

La regione di legame al DNA è una diretta estensione della regione di dimerizzazione, le due α-eliche divergono dalla struttura superavvolta e contattano il solco maggiore del DNA in direzioni opposte e ognuna lega metà della sequenza regolatrice (Luscombe et al., 2000).

Sono state identificate 17 proteine in *Neurospra crassa* appartenenti a questa classe (Borkovich et al., 2004), 53 in uomo (Vinson et al., 2002), 17 in *S. cerevisiae* e 27 in *Drosophila* (Fassler et al., 2002). Tra i fattori bZIP caratterizzati in funghi, GCN4 di *Saccharomyces cerevisiae* e *Candida albicans* (Martchenko et al., 2007) rappresenta un attivatore trascrizionale per geni coinvolti nella biosintesi degli aminoacidi. In presenza di stress ambientali, quali la carenza aminoacidica, viene indotta la traduzione di GCN4. Esso si lega come omodimero alla sequenza TGACTC localizzata a monte di molti geni indotti in condizioni di carenza di aminoacidi.

Alla seconda classe appartengono i domini elica-giro-elica che sono propri solo degli eucarioti. Questi motivi sono costituiti da domini bipartiti conservati coinvolti nel legame al DNA e nella dimerizzazione (Borkovich et al., 2004).

La regione basica responsabile del legame al DNA si trova all’N-terminale della proteina, è costituita da 15 amminoacidi e riconosce una sequenza consenso di sei nucleotidi 5’-CANNTG-3’ detta E-box. Diverse famiglie di bHLH riconoscono diversi E-box ed è proprio questa eterogeneità insieme alla formazione di dimeri con diverse proteine che determina come questi fattori di trascrizione vadano a regolare diversi processi all’interno della cellula (Jones, 2004).

Inoltre è stato dimostrato che le regioni fiancheggianti la sequenza consenso possono ricoprire un ruolo nel determinare la specificità di legame e che residui presenti nel loop possono riconoscere queste sequenze esterne all'E-box (Toledo-ortiz et al., 2003).
Al C-terminale c'è il dominio HLH che differisce dai bZIP poiché è formato da due α-eliche separate da un loop che dimerizzano formando un fascio di quattro eliche. Come per i bZIP le eliche responsabili della dimerizzazione interagiscono formando una struttura superavvolta mentre l'elica responsabile del legame con il DNA si inserisce nel solco maggiore. La separazione dei due domini permette una maggiore flessibilità nel posizionamento dell'elica del DBD nel sito di riconoscimento presente nel DNA (Luscombe et al., 2000).
Un sottogruppo dei bHLH sono i fattori di trascrizione APSES che sono tipici dei funghi e regolano processi morfogenetici e la differenziazione negli ascomiceti (Shelest, 2008).
La seconda superfamiglia è costituita dai domini "*zinc finger*" che rappresentano il gruppo più abbondante dei fattori di trascrizione presenti nei genomi eucariotici. Il loro dominio di legame al DNA è caratterizzato dalla coordinazione di uno o due atomi di zinco da parte di residui di cisteina o istidina. Questi ioni metallici conferiscono una ottima stabilità strutturale alla proteina e sono indispensabili per il corretto *folding* del dominio di legame al DNA (Luscombe et al., 2000).
Gli *zinc finger* a loro volta sono suddivisi in tre gruppi.
Il primo gruppo è formato da quei domini in cui l'atomo di zinco è coordinato da quattro residui di cisteine (Cys2-Cys2 *type*). Il dominio di legame al DNA di questi fattori è costituito da due foglietti β irregolari e antiparalleli, da una α elica e da un loop esteso. Queste proteine interagiscono con sequenze di DNA che contengono la sequenza target A/TGATAA/G e sono denominati fattori GATA. L'elica e i loop che connettono i foglietti β legano il DNA nel solco maggiore ma si creano anche interazioni con i gruppi fosfato del DNA. I fattori GATA sono presenti nei funghi, piante e metazoi dove possono agire sia come attivatori che come repressori. Alcuni di questi fattori di trascrizione in *Neurospora* e in altri funghi sono coinvolti nell'espressione di geni strutturali coinvolti nell'assimilazione del nitrato e attivano l'espressione di nitrato e nitrito reduttasi (Scazzocchio, 2000).
Il secondo gruppo è costituito dai domini in cui due atomi di zinco sono coordinati da sei residui di cisteina (Cys6 *type* o *zinc cluster*). Questo dominio di legame, tipico dei funghi, è stato caratterizzato per la prima volta nella proteina GAL4 di *S. cerevisiae* che è responsabile della regolazione di geni coinvolti nel catabolismo del galattosio. Il

motivo è costituito da sei cisteine che si trovano su due α eliche separate da un loop e che coordinano due atomi di zinco. Il domino di legame al DNA è formato da trinucleotidi, conservati e disposti in una configurazione simmetrica, che sono spaziati da una sequenza variabile ma di lunghezza definita (Borkovich et al., 2004).

I fattori di trascrizione appartenenti a questo gruppo possono legarsi al DNA come monomeri, omodimeri o eterodimeri. La struttura proteica di questi fattori di trascrizione può essere suddivisa in tre domini: di legame al DNA, regolativo e di attivazione. Il DBD è a sua volta suddiviso in tre regioni: *zinc finger*, linker, dominio di dimerizzazione. All' N-terminale della proteina si trova lo *zinc finger* che è suddiviso in 2 sottostrutture ognuna formata da 3 cisteine circondate da residui basici e separate da un loop e insieme formano una struttura di due corte α-eliche tra le quali sono presenti i due atomi di zinco che si legano alle sei cisteine. Nel loop c'è un residuo di prolina che conferisce flessibilità al dominio. Quando formano omodimeri riconoscono sul DNA una coppia di triplette CGG interagendo con il solco maggiore. La regione linker è posizionata al C-terminale del DBD e non possiede caratteristiche conservate ma si pensa che alcune sue regioni contribuiscano a determinare la specificità di legame. Il domino di dimerizzazione è al C-terminale del linker ed è caratterizzato da ripetizioni di sette residui che determinano una struttura superavvolta responsabile dell'interazione tra proteine. La regione di regolazione si trova tra il DBD e il dominio di attivazione ma non è presente in tutti i fattori di trascrizione, è lunga 80 amminoacidi ed è coinvolta nella regolazione dell'attività trascrizionale. Ci sono casi in cui la delezione di questa regione attiva il TF in modo costitutivo. Il dominio di attivazione è un dominio acido e non è conservato. Da alcuni studi è emerso che molti di questi fattori di trascrizione riconoscono sequenze regolatrici del DNA caratterizate dalla presenza di una tripletta CGG singola o ripetuta in modo simmetrico o asimmetrico anche se alcune possono subire modificazioni. Molti fattori contribuiscono alla specificità di legame tra cui i componenti del DBD ma anche i nucleotidi che fiancheggiano la tripletta ed è fondamentale anche l'orientamento della sequenza CGG e lo spazio presente tra le ripetizioni. Anche se la maggior parte degli *zinc cluster* si legano sottoforma di omodimeri ci sono anche molti casi in cui funzionano come monomeri o come eterodimeri che possono essere formati con TF appartenenti alla stessa famiglia ma anche con membri di altre famiglie. Questi fattori di trascrizione possono essere suddivisi in base alla loro iniziale localizzazione cellulare: alcuni sono sempre nel nucleo mentre altri prima sono localizzati nel citoplasma e solo in un secondo momento vengono attivati e traslocati nel nucleo. La maggior parte di quelli permanentemente

presenti nel nucleo si trovano già legati al promotore ma sono in forma inattiva e svolgono la loro azione solo a seguito di attivazione. Ci sono inoltre casi in cui lo *zinc cluster* necessita il reclutamento di cofattori trascrizionali o di proteine coinvolte nella modificazione della cromatina o degli istoni per facilitare la trascrizione (MacPherson et al., 2006).

Questi fattori sono coinvolti in diversi processi di metabolismo primario e secondario ad esempio in *Neurospora* sono coinvolti nel metabolismo dell'azoto, in *S. cerevisiae* alcuni di questi fattori sono coinvolti nel metabolismo del galattosio mentre in *A. nidulans* sono coinvolti nell'utilizzazione dell'acetato (Borkovich et al., 2004).

Ci sono anche casi in cui a valle degli *zinc cluster* è presente anche un secondo dominio di legame al DNA tipico dei funghi che è chiamato *"fungal specific transcription factor"* il cui ruolo è ancora sconosciuto ma si trova in fattori di trascrizione che regolano il metabolismo degli zuccheri o degli amminoacidi, la gluconeogenesi, la respirazione o il catabolismo degli acidi grassi (Shelest, 2008).

Il terzo gruppo è formato dai motivi in cui l'atomo di zinco è coordinato da due residui di cisteina e da due residui di istidina (Cys2-His2 *type*). Questo gruppo è predominante negli eucarioti e il primo fattore di trascrizione identificato fu TFIIIA di *Xenopus*. Ogni *zinc finger* comprende solo 30 amminoacidi e si ripiega in un motivo beta-beta-alfa stabilizzato dal coordinamento dell'atomo di zinco con due cisteine presenti nel foglietto β e con due istidine presenti nell'α-elica. Solitamente nella proteina sono presenti più *zinc finger* che contattano il DNA in più punti. La regione linker che separa due domini *zinc finger* sulla stessa proteina è un importante elemento strutturale che serve per controllare lo spazio dei due domini sul DNA. Il linker nella maggior parte dei casi è composto da cinque residui tra l'ultima istidina di un dominio e la primo amminoacido aromatico conservato presente nel secondo. Un buon numero di fattori di trascrizione appartenenti a questo gruppo sono coinvolti nella regolazione del ciclo cellulare (Wolfe et al., 2000).

Alla terza superfamiglia appartengono i domini elica-giro-elica caratterizzati da domini con strutture α-elica impaccate insieme in una struttura superavvolta. Il dominio α-elica più semplice e frequente consiste di quattro α-eliche disposte in un fascio con l'asse di ognuna pressoché parallelo a quelli delle altre ma il numero di eliche presenti nelle diverse strutture può variare da tre a sei. La seconda α-elica, detta elica di riconoscimento si lega al DNA nel solco maggiore, si creano però anche dei contatti tra la prima elica del motivo e il *backbone* del DNA (Luscombe et al., 2000). All'interno di questa superfamiglia ritroviamo numerose classi diverse di fattori di trascrizione.

Una classe è rappresentata dagli *homeodomain* che sono domini di legame al DNA di circa 60 amminoacidi codificati da geni detti *homeobox*. È stato dimostrato che sono coinvolti nella regolazione di processi di sviluppo e differenziazione. La loro struttura è costituita da tre α-eliche collegate da loop, le eliche 2 e 3 fanno parte del motivo elica-giro-elica mentre l'elica 1 è connessa tramite un loop alla seconda elica. L'α-elica 1 è l'elica di riconoscimento ed ha un diverso orientamento rispetto alle altre. L'elica di riconoscimento è posizionata nel solco maggiore e il complesso DNA-proteina è stabilizzato anche da interazioni del peptide con il *backbone* (Yang, 1998).
Una classe è rappresentata dai domini *Forkhead/Winged helix* che hanno una struttura α/β formata da 2 loop detti W1 e W2, tre α-eliche e tre foglietti β. L'N-terminale della proteina comprende le eliche mentre al C-terminale due foglietti formano una struttura a filamento β antiparallelo. Tra questi due foglietti si trova il loop W1 mentre il W2 si estende alla fine del dominio dopo l'ultimo foglietto β. Anche in questo caso l'elica 3 è quella di riconoscimento e contatta il DNA nel solco maggiore mentre anche il loop W2 contatta il DNA ma nel solco minore (Gajiwala and Burley, 2000).
Una sottoclasse appartenente alla superfamilgia elica-giro-elica è rappresentata dai fattori di trascrizione che posseggono un dominio MYB che è formato da ripetizioni di moduli elica-giro-elica di 53 amminoacidi ciascuno. Queste ripetizioni contengono residui di triptofano distribuiti in modo regolare che nel *folding* della proteina formano un cluster. Queste proteine sono presenti in tutti gli eucarioti ma sono particolarmente abbondanti nelle piante dove sono coinvolti nella regolazione di *pathway* biosintetici per la produzione di fenilpropanoidi e triptofano ma agiscono anche nel nella regolazione del ciclo cellulare e nella regolazione dell'espressione influenzata dal ritmo circadiano (Zimmermann et al., 2004).
Un'altra classe è quella degli *heat shock factors* HSF che sono fattori di trascrizione coinvolti nella regolazione di proteine attivate da *shock* termici. Questi fattori di trascrizione formano omo-tetrameri che sono localizzati nel citoplasma e solo a seguito dello stress vengono attivati e importati nel nucleo dove si legano ai promotori dei geni regolati da stress termici. Le sequenze regolatrici a cui si legano sono caratterizzate dalla presenza di un motivo composto da ripetizioni invertite del modulo nGAAn. Gli HSF sono composti da un dominio di legame al DNA di tipo elica-giro-elica e un dominio di trimerizzazione (Hardy and Nelson, 2000).
Un'ulteriore dominio appartenente a questa superfamiglia è rappresentato dal dominio TEA/ATSS che è presente all'N-terminale della proteina ed è composto da 66-76 amminoacidi. Questo dominio si trova in fattori di trascrizione che sono coinvolti nello

sviluppo. Un esempio è la proteina presente nei funghi AbaA che regola la formazione delle spore aploidi (Schweizer et al., 2000).

Alla quarta superfamiglia appartengono i domini β scaffold che legano il DNA nel solco minore. È però necessario sottolineare che diventa difficile trovare caratteristiche comuni per tutti i fattori di trascrizione in questa superfamiglia e alcuni dei gruppi presenti non posseggono la caratteristica di β scaffold.

I MADS box sono fattori di trascrizione appartenenti a tale gruppo e sono domini conservati di 56 amminoacidi, la regione N-terminale è idrofilica e presenta un buon numero di residui basici mentre la regione C-terminale è idrofobica ed è necessaria per la dimerizzazione. Molti fattori MADS box formano complessi di regolazione andando a reclutare altri fattori di trascrizione. A questa classe appartiene la proteina MCM1 di *S. cerevisiae* che gioca un ruolo centrale nel formare complessi di attivazione o repressione per la determinazione del tipo cellulare a o α del lievito e nel controllo del processo di mating (Shore and Sharrocks, 1995).

Un altro gruppo è rappresentato dai domini *High mobility group* (HMG) che hanno una struttura formata da 3 eliche raggruppate in modo irregolare. Questo domino può essere presente in singola copia oppure ripetuto ed è presente in proteine coinvolte oltre che nella trascrizione anche nella replicazione, riparazione dei filamenti, tutti processi per cui è previsto una curvatura e svolgimento della cromatina. Le proteine che posseggono questo dominio si dividono in due categorie in base al riconoscimento specifico o aspecifico del DNA. Generalmente le prime posseggono un solo dominio HMG mentre le altre ne contengono più ripetizioni.

Ad un altro gruppo appartengono i domini *histone fold* che consistono di una lunga α-elica fiancheggiata ad ogni lato da una corta α-elica e assumono una struttura che prende il nome di *histone fold.* Si suppone che questa struttura sia derivante da una duplicazione genica del motivo elica-strand-elica dove la fusione di due eliche ha portato alla creazione della lunga elica centrale (Burley et al., 1997).

Alla superfamiglia 0 appartengono i fattori che contengono altre caratteristiche e sono presenti fattori di trascrizione, come ACE1 di *S. cerevisiae* e AMT1 di *Candida glabrata*, coinvolti nel metabolismo del rame mediante l'attivazione di geni che codificano per metallotioneine e caratterizzati dalla presenza di un dominio di legame al DNA ricco in cisteine. Altro fattore molto simile é MAC1 di *S. cerevisiae* che viene richiesto per la regolazione della trascrizione di geni coinvolti nell'utilizzo Fe/Cu e nella risposta allo stress.

A questa superfamiglia appartengono anche fattori di trascrizione tipici delle pianta come le proteine ERF che hanno un dominio di legame al DNA di 58-59 amminoacidi con una struttura formata da 3 foglietti β. Queste proteine si legano a motivi GCC presenti nei promotori di geni coinvolti nella risposta a patogeni e a motivi CRT che presentano ripetizioni di C e si trovano a monte di geni attivati in condizioni di disidratazione o in risposta a basse temperature. Queste proteine sono responsabili della risposta a numerosi stress abiotici e sono responsabili dell'induzione di geni in presenza di etilene (Singh, 2002).
Anche il domino NAC è tipico delle piante ed ha una struttura a foglietto β antiparallelo circondato da due α eliche poste alle estremità. Costituisce una delle più abbondanti famiglie di fattori di trascrizioni unici di pianta e sono coinvolti nella regolazione dello sviluppo embrionale, vegetativo e dei fiori, nella formazione delle radici laterali, nella risposta a stress abiotici e biotici come ad esempio l'infezione di patogeni (Olsen et al., 2005).

2.1 Transcriptional activator trap

A differenza dei domini di legame, che sono stati ampiamente caratterizzati sia funzionalmente che strutturalmente, i domini di attivazione (AD) non condividono motivi di struttura facilmente riconoscibili. La particolare componibilità dei fattori di trascrizione ha fornito però gli strumenti per la loro individuazione.
Nel 1987, Ptashne e collaboratori hanno cercato di identificare quali fossero le caratteristiche che permettono di attribuire ad una proteina il ruolo di dominio di attivazione generando fattori di trascrizione artificiali codificati da frammenti del DNA genomico di *Escherichia coli* fusi con il DBD di GAL4 di lievito. Tutti gli attivatori individuati erano costituiti da residui acidi e non mostravano omologia di sequenza quando venivano comparati tra loro o con le sequenze AD di GCN4 e GAL4. Inoltre la maggior parte delle sequenze attivatrici conteneva β-turn aventi residui acidi affiancati da α-eliche composte da aminoacidi idrofobici e corti segmenti che formano α-eliche anfifiliche (Ma and Ptashne, 1987).
Due studi successivi hanno sviluppato un efficiente sistema in lievito per l'isolamento di geni di riso (Ye et al., 2004) e di lievito stesso (Titz et al., 2006) codificanti fattori trascrizionali e proteine con funzione di attivatori. Entrambi sono basati su una variazione del sistema del doppio ibrido in cui i cDNA del genoma di riso e le ORF di lievito rispettivamente sono stati fusi con il dominio di legame al DNA di GAL4 mentre a monte dei geni *reporter* era posta la UAS di GAL4. La proteina di fusione al Gal4-DBD

che è in grado di attivare la trascrizione del gene *reporter* è detta auto-attivatore e quindi possiede le proprietà di un domino di attivazione (Titz et al., 2006). Con questo *screening* non solo si identificavano i possibili attivatori ma si valutava anche il livello di espressione utilizzando saggi che permettevano di misurare l'espressione dei due geni *reporter HIS3* e *LacZ*. Da questo studio emerse che l'attività trascrizionale è strettamente correlata ad un forte livello di espressione dei geni *reporter* e inoltre con questo *screening* si riuscirono ad identificare alcuni fattori di trascrizione non ancora catalogati come tali in YPD database e questo suggeriva che ci fossero molte altre proteine con funzione di attivatori trascrizionali ancora sconosciute. Inoltre l'attività risulta anche essere correlata alla localizzazione nucleare, le proteine appartenenti alla categoria degli attivatori forti hanno un maggior numero di proteine con localizzazione nucleare rispetto alle attivazioni più deboli (Titz et al., 2006).

È noto da tempo che alcuni amminoacidi sono più abbondanti nella sequenza dei fattori di trascrizione in particolare residui acidi e basici e certi altri amminoacidi come la glutammina. *Cluster* di amminoacidi basici agiscono anche da segnale di localizzazione nucleare.

Dalle proprietà chimico-fisiche è emerso che i fattori di trascrizione posseggono un basso punto isoelettrico, una bassa idrofobicità e un alto peso molecolare. Si possono distinguere diverse tipologie di attivatori: acidi, ricchi in prolina, in serine o in glutammine (Titz et al., 2006).

Le caratteristiche che deve possedere un potenziale nuovo fattore di trascrizione sono:

- attivazione forte del gene *reporter*
- localizzazione nucleare
- interazioni con proteine nucleari.

Dallo studio condotto in riso (Ye et al., 2004) è emerso che alcuni degli attivatori trascrizionali identificati mancavano di un dominio di legame al DNA definito e questo può significare che queste proteine svolgono ruoli di co-attivazione. In questo studio sono state isolate anche proteine ribosomali suggerendo che tali proteine possano avere un ruolo addizionale coinvolto nella attivazione della trascrizione ma anche chinasi, proteine coinvolte nella riparazione del DNA, nel rimodellamento della cromatina, proteine a funzione sconosciuta e anche proteine note possedere altre funzioni non correlate alla trascrizione come ad esempio enzimi metabolici.

Nello studio condotto in lievito emerge l'incapacità dello *screening* di identificare tutti i regolatori della trascrizione; solo il 40% degli attivatori che sono predetti avere tale funzione nel database YPD sono identificati in questo studio. Questo risultato

suggerisce una possibile percentuale del 60% di falsi negativi. Questa discrepanza può essere dovuta al fatto che alcune proteine annotate come regolatori della trascrizione non agiscano in tal modo in questo saggio, altre proteine possono richiedere specifici promotori, cofattori o condizioni di crescita particolari per poter svolgere la loro funzione. Ad esempio in lievito alcuni attivatori sono attivi solo in condizioni di assenza di glucosio nel terreno. Inoltre lo stesso Gal4 ha una bassa attivazione se fuso ad un altro Gal4-DBD, ad esempio sembra che il DBD possa in qualche modo inibire la capacità di attivare la trascrizione di alcune proteine (Titz et al., 2006).

Questi lavori hanno fornito dei buoni esempi di strategie di identificazione di attivatori trascrizionali utilizzabili in altri organismi eucarioti come i funghi e, come in questo caso, in *T. melanosporum, L.bicolor e P. trichocarpa*.

Non bisogna però sottovalutare la presenza di falsi positivi che possono essere dovuti a motivi tecnici o metodologici. I motivi tecnici sono principalmente a carico di mutazioni spontanee insorte nel lievito che portano accidentalmente all'espressione di un gene *reporter* nello *screening;* non includono quindi i cloni selezionati più volte e che sono in grado di attivare più di un gene *reporter*.

Quelli metodologici sono dovuti al tipo di vettore utilizzato per la costruzione della genoteca poiché otteniamo una proteina di fusione con il DNA binding domain di GAL4 che contiene un segnale di localizzazione nucleare. La proteina di fusione risulta essere espressa in modo forzato nel nucleo e può in tal modo risultare attivatore della trascrizione perché possiede una regione acida ma in realtà non essendo localizzata nel nucleo non possiede tale attività.

Le proteine non convenzionali identificate dallo *screening* potrebbero però non rappresentare dei falsi positivi ma rientrare nella categoria delle *moonlighting protein* cioè di proteine multifunzionali per cui non è ancora stata identificata l'attività di regolatore della trascrizione. Diventa di conseguenza fondamentale capire se le proteine codificate da queste sequenze, da noi isolate, entrano nel nucleo per la presenza di un NLS e quindi svolgono funzione di attivatore trascrizionale o se sono dei falsi positivi.

3. Le proteine nucleari

Le proteine nucleari svolgono un importante ruolo nella trascrizione, nel processamento dell'mRNA, nella replicazione e nell'organizzazione dei cromosomi. Molte proteine nucleari hanno un segnale di localizzazione nucleare intrinseco (NLS) che media il trasporto attivo attraverso il complesso del poro nucleare. Il segnale di localizzazione meglio caratterizzato è formato da una corta sequenza ricca di residui basici che viene riconosciuta dalla proteina Kap o dall'eterodimero di importine α/β (Ueki et al., 1998).

Il passaggio delle proteine dentro e fuori il nucleo avviene attraverso una struttura proteica detta complesso del poro nucleare. Il poro nucleare permette la diffusione passiva di piccole molecole comprese tra i 50-60kDa mentre il passaggio di molecole più grandi viene regolato e richiede che la proteina comprenda nella sua sequenza un motivo di localizzazione nucleare (NLS) che ne permetta l'interazione con le proteine responsabili della traslocazione nel nucleo. Le proteine che costituiscono il poro nella sua struttura centrale sono chiamate nucleoporine. Il trasporto attivo di macromolecole è mediato da proteine solubili che agiscono da carrier che sono definite carioferine e in particolare quelle che mediano l'importo appartengono alla famiglia delle carioferine β che sono caratterizzate dalla capacità di legare le nucleporine e di formare un complesso con la Ran-GTP. L'importina β è in grado di legare la proteina cargo direttamente oppure tramite una proteina adattatrice che è l'importina α. Quest'ultima proteina è la responsabile del riconoscimento dell'NLS presente nelle proteine che devono essere trasportate nel nucleo. La regione C-terminale dell'importina α interagisce con la proteina cargo mentre la regione N-terminale è responsabile dell'interazione con l'importina β tramite un dominio detto regione di autoinibizione caratterizzato dalla presenza di residui basici e importante per il successivo rilascio del cargo nel nucleo. L'importina non interagisce direttamente con le nucleoporine (Macara, 2001).

Affinché ci sia un efficiente trasporto attivo devono essere rispettate alcune condizioni: le proteine che legano e trasportano il cargo devono essere presenti solo in un compartimento cellulare, il cargo può essere modificato in uno dei due compartimenti in modo che non interagisca più con il carrier, il legame carrier-cargo può essere regolato da un secondo fattore in modo che l'assemblaggio avvenga in un compartimento mentre il rilascio avvenga nell'altro.

Nel meccanismo classico si crea una distribuzione asimmetrica della Ran per cui all'interno del nucleo è più abbondante nella forma associata a GTP (Ran-GTP) mentre all'esterno è più abbondante quella associata a GDP (Ran-GDP). Questo gradiente è assicurato dalla localizzazione delle proteine che modificano lo stato della Ran per cui la Ran-GEF (*Guanilic nucleotide Exchange Factor*) è costitutivamente nel nucleo mentre la Ran-GAP (*GTPase Activator Protein*) è nel citoplasma. Nel nucleo l'importina β si associa alla Ran-GTP staccandosi dal complesso dell'importina α e della proteina cargo e una volta legata a tale fattore viene ritraslocata nel citoplasma. L'importina α nella forma non legata alla β espone la sua regione di autoinibizione che causa il rilascio della proteina trasportata. L'importina α in forma libera interagisce con una proteina di trasporto e con la Ran-GTP e ritorna nel citoplasma. Una volta nel citoplasma questi complessi si dissociano e il GTP viene idrolizzato in GDP dalla Ran-GAP, il trasporto di una macromolecola nel nucleo richiede quindi due molecole di GTP. La forma Ran-GDP viene poi riconosciuta da un trasportatore che la trasloca nel nucleo per diffusione passiva dove la Ran-GEF causa la dissociazione del complesso e converte la Ran al suo stato di legame al GTP (Macara, 2001).

Il segnale di localizzazione nucleare può essere caratterizzato da uno o due zone contenenti residui basici e quindi essere definito rispettivamente monopartito o bipartito. Un esempio di NLS monopartito è quello dell'antigene di SV40 che possiede la sequenza PKKKRRV mentre la sequenza KRPAATKKAGQAKKKK della nucleoplasmina è un esempio di motivo bipartito.

Si assume che questo sia il *pathway* maggiormente utilizzato dalla cellula ma non si esclude la presenza di altri meccanismi per la traslocazione nel nucleo di macromolecole. Ad esempio studiando il proteoma di *S. cerevisiae* si è notato che mediante analisi in silico delle 5850 proteine di lievito il 45% mostrava la presenza di un NLS monopartito o bipartito. Mentre mediante analisi in vivo attraverso la fusione delle proteine con la GFP solo il 25,8% risultava nel nucleo e di queste solo il 57% possedeva un NLS predicibile con programmi bioinformatici. Si può dedurre che proteine nucleare prive di questo NLS canonico possano entrare nel nucleo sfruttando altri meccanismi di trasporto (Lange et al., 2007).

L'NLS non è facilmente predicibile attraverso analisi bioinformatiche poiché il motivo non è chiaro e ben definito, inoltre anche se si identifica questo motivo *in silico* non si riesce a determinare se nella cellula è funzionale (Rhee et al., 2000).

Nell'analisi di proteine a localizzazione nucleare ci si trova in presenza di diverse condizioni: la proteina può non avere un NLS riconoscibile oppure se lo possiede

potrebbe essere inattivo, esposto e quindi riconosciuto dalle proteine di trasporto oppure nascosto ed esposto solo a seguito di eventi che causano una modificazione della proteina come ad esempio la fosforilazione. Ci sono inoltre anche casi in cui la proteina subisce delle modificazioni che ne inibiscono l'interazione con le proteine di trasporto (Macara, 2001).

Ci possono essere diversi fattori che rendono inattivo un NLS ad esempio può essere troppo debole il legame con l'importina α e quindi la proteina non viene legata oppure il legame è troppo forte per cui una volta nel nucleo non si dissociano nonostante la presenza di altri fattori che dovrebbero favorire il distacco del cargo. L'efficienza di importo dipende dalla formazione del complesso di traslocazione; quindi fattori decisivi sono l'affinità di legame tra la proteina e l'importina α e la concentrazione dell'importina α (Lange et al., 2007).

Alternativamente una proteina può non entrare nel nucleo perché possiede alta affinità per una proteina citoplasmatica che la ancora in questo comparto cellulare (Cyert, 2001).

Per tutti questi motivi un segnale di localizzazione nucleare per essere definito tale deve rispettare alcuni criteri:

- La sequenza deve essere necessaria per l'importo, quindi se mutata si deve notare un drastico calo della traslocazione di tale proteina nel nucleo.
- La sequenza deve essere in grado da sola di mediare l'importo di una proteina non presente nel nucleo.
- L'interazione della proteina di interesse con i recettori che ne mediano l'importo deve essere dipendente dalla presenza di questa sequenza.
- Per poter definire quale meccanismo utilizza per entrare nel nucleo è necessario dimostrare che impedendo questo tipo di trasporto è bloccato anche l'importo della proteina.

(Lange et al., 2007)

3.1 Nuclear transportation trap

È stato sviluppato un metodo che permette di saggiare se la proteina che stiamo studiando possiede un segnale di localizzazione nucleare. La strategia è basata sull'espressione in lievito di una tripla proteina di fusione che contiene la sequenza batterica LexA modificata, il dominio di attivazione Gal4AD e il cDNA della proteina da saggiare clonata in frame a valle di Gal4AD. La sequenza LexA è stata modificata in quanto si è notato che, nonostante fosse di derivazione batterica, possedeva al suo

interno una sequenza, non identificata in precedenza, che agiva come segnale di localizzazione nucleare. Il vettore così costruito è denominato pNIA. Se la sequenza che inseriamo nel vettore possiede un NLS il prodotto di fusione entrerà nel nucleo di lievito e attiverà la trascrizione dei geni *reporter* messi a monte del sito operatore LexA nel ceppo di lievito L40. I geni *reporter* sono *LacZ* e *HIS3* di conseguenza la proteina in grado di traslocare nel nucleo mostrerà l'attivazione della β-galattosidasi e la capacità di crescita su terreno privo di istidina e addizionato di 3-aminotriazolo (3AT). È stato notato però che in assenza della proteina da saggiare la proteina di fusione LexA-Gal4AD può diffondere nel nucleo a causa delle sue dimensioni ridotte e quindi attivare ugualmente l'espressione dei geni *reporter* (Rhee et al., 2000).
Per questo motivo in uno studio successivo il vettore pNIA è stato modificato per eliminare il *back-ground* di crescita del vettore vuoto in condizioni selettive. Questo *back-ground* è causato per due motivi: il primo è legato al fatto che il vettore è multicopia e quindi si ottiene una sovraespressione della proteina in esame, il secondo è legato al fatto che la proteina di fusione che si crea è di piccole dimensioni e può passare nel nucleo per diffusione passiva. Per risolvere il primo problema Marshall e collaboratori hanno deciso di trasferire la cassetta di espressione in un vettore a singola copia mentre per eliminare la diffusione passiva nel nucleo è stato inserito nel vettore un frammento corrispondente ad una porzione della *maltose binding protein* di *E.coli*, posto in frame tra LexA-DBD e Gal4-AD in modo da ottenere un aumento del peso molecolare della proteina di fusione sino a 80kDa. Questo nuovo vettore, pNIA-CEN-MBP, non evidenzia nessun *back-ground* (Marshall et al., 2007).

4. *Moonlighting protein*

Il concetto che un gene codifica per una proteina in grado di svolger un'unica funzione è ormai da tempo abbandonato poiché è risaputo che una proteina può svolgere diverse funzioni. Un esempio di proteine multifunzionali è rappresentato dalla categoria delle *moonlighting protein* (Jeffery, 1999).
Il termine *moonlighting* significa letteralmente *"to do [paid] work, [usually at night], in addition to one's regular employment"* ed è stato utilizzato per la prima volta da Jeffery per definire tutte quelle proteine che posseggono una seconda funzione oltre a quella già nota. Inoltre questo ruolo ulteriore spesso non ha nessun collegamento con il ruolo metabolico. Si ritrovano, infatti, enzimi che possono agire come fattori di trascrizione, partecipare nell'assemblaggio o degradazione degli organelli, contribuire alla stabilità

del genoma mitocondriale o essere coinvolti nei meccanismi di riparazione e sintesi del DNA (Gancedo and Flores, 2008).

È importante sottolineare che non si devono definire *moonlighting*: le proteine risultanti da fusioni geniche, proteine omologhe ma non identiche, varianti di splicing, proteine che possono subire diverse modificazioni post-traduzionali, e proteine che svolgono la stessa funzione però in diversi comparti cellulari (Jeffery, 1999).

Le *moonlighting protein* sono definite multifunzionali poiché svolgono più funzioni indipendenti e spesso non correlate tra loro, ciò significa che se viene inattivata una funzione attraverso mutazione questa non influenza gli altri ruoli e viceversa (Huberts and van der Klei, 2010).

La funzione della proteina può variare in base a diversi fattori:

- diversa localizzazione: la medesima proteina possiede diverse funzioni in base alla sua localizzazione. Ad esempio la proteina batterica PutA funziona da deidrogenasi se è ancorata alla membrana mentre se è libera nel citoplasma agisce come repressore trascrizionale. Oppure ci sono proteine che hanno un diverso ruolo se secrete dalla cellula come la fosfoglucosio isomerasi che nel citoplasma converte il glucosio 6-fosfato in fruttosio 6-fosfato mentre se è secreta possiede quattro funzioni tra cui azione di neuroleuchina e di fattore di crescita neuronale.
- Espressione differenziale: in base al tipo cellulare in cui è espressa svolge diverse funzioni.
- Oligomerizzazione: l'attività cambia a seconda che la proteina sia un monomero o un multimero.
- Concentrazione del ligando o del substrato: la funzione può variare in base alla disponibilità di cofattori, substrati o ligandi. Anche in tal caso si può prendere come esempio la proteina PutA che si lega alla membrana se il suo substrato è presente ad alte concentrazioni mentre lega il DNA quando diminuisce il substrato.
- Siti multipli di legame: le funzioni dipendono dalla presenza nella protiena di diversi siti di legame.
- Formazione di complessi: la proteina modifica la sua funzione poiché si lega ad altri peptidi che fanno parte di subunità che formano un complesso proteico. (Jeffery, 1999)

Alle *moonlighting protein* possono appartenere diverse categorie di proteine come ad esempio enzimi e fattori di trascrizione (Jeffery, 2009).

Sembra plausibile poter affermare che il ruolo primario della proteina sia quello enzimatico poiché i siti richiesti per l'attività catalitica sono maggiormente definiti e occupano una piccola porzione della proteina rendendo la parte della struttura non richiesta per l'attività metabolica più flessibile e quindi maggiormente tollerante a cambiamenti (Gancedo and Flores, 2008).

Ci sono numerosi esempi di proteine che sono state scoperte possedere diverse funzioni all'interno della cellula. Nel nostro caso siamo particolarmente interessati a tutte quelle che sono state scoperte essere coinvolte nella trascrizione genica dato che con l'esperimento condotto per l'identificazione degli attivatori trascrizionali si isolano anche delle proteine non convenzionali. In letteratura si trovano esempi sia di enzimi che legano direttamente il DNA sia proteine che regolano la trascrizione senza una diretta interazione con il DNA ma attraverso intreazioni di tipo proteina-proteina.

La proteina hRoDH-E2 nel reticolo endoplasmatico funziona come deidrogenasi/reduttasi di retinoidi e idrossi-steroidi mentre nel nucleo agisce come repressore della trascrizione. La proteina viene reclutata a livello dei promotori della profilaggrina dove influenza l'attività del fattore di trascrizione Sp1, hRoDH-E2 non si lega direttamente al DNA ma agisce modificando la stato di acetilazione e deacetilazione degli istoni vicino al sito di inizio della trascrizione. Il suo ruolo nella repressione non coinvolge l'attività catalitica ma è dovuto a un meccanismo di interazione tra proteine (Markova et al., 2006).

La galattochinasi Gal1 di *Kluyveromyces lactis* è responsabile della fosforilazione del galattosio a galattosio 1-fosfato ma è anche in grado di attivare la trascrizione dei geni *GAL.* Gal1 interagisce e blocca l'attività del repressore della trascrizione Gal80 che in tal modo non riesce più a inibire l'attività del fattore di trascrizione Gal4 che così è in grado di promuove l'espressione dei geni responsabili per l'utilizzo del galattosio come fonte di carbonio (Gancedo and Flores, 2008).

La malato deidrogenasi MDH1 è un enzima metabolico che converte il malato in ossaloacetato ma è presente anche nel nucleo dove lega il fattore di trascrizione p53 che è un oncosoppressore coinvolto nella regolazione, in base alla risposta a stress, dell'apoptosi, della riparazione del DNA e del mantenimento della stabilità del genoma. In condizioni di carenza di glucosio MDH1, che trasloca nel nucleo in modo autonomo da p53, lega il sito di legame al DNA di p53 andando a stabilizzare il fattore di trascrizione e impedendone l'esporto dal nucleo. Si è anche dimostrato che MDH1 ha un ruolo come co-attivatore di p53 e che in assenza di glucosio le due proteine interagiscono e legano i promotori dei geni regolati da p53. Questo suggerisce una

comunicazione tra il metabolismo energetico e l'azione di p53, in particolare MDH1 è in grado di regolare l'apoptosi in condizioni di privazione di glucosio (Lee et al., 2009).
L'omocitrato sintasi LYS20 è un enzima metabolico coinvolto nella biosintesi della lisina ed è l'unico enzima di questo *pathway* che oltre ad essere presente nel citoplasma e nei mitocondri si trova anche nel nucleo. In particolare LYS20 è funzionalmente correlata con la proteina ESA1, una istone acetilasi, e questa interazione sembra essere mediata dal gene *HTZ1* che codifica per la variante istonica H2A.Z. Si pensa che LYS20 sia coinvolta nel processo di riparazione del DNA e che possa avere una attività acetilasica sull'istone H4 e che la sua attività sia ristretta alla presenza nel nucleosoma dell'istone H2A.Z (Scott and Pillus, 2010).
In numerosi studi è emerso come gli enzimi glicolitici siano in grado di svolgere altre funzioni oltre a quella metabolica. Si stima che 7 sui 10 enzimi coinvolti nella glicolisi appartengano alla categoria *moonlighting*.
Un esempio è rappresentato dalla esochinasi2 che agisce sia come enzima, fosforila il glucosio in glucosio 6-fosfato, sia come repressore della trascrizione. Ci sono evidenze che la localizzazione nucleare della esochinasi2 dipenda dall'interazione con il repressore trascrizionale Mig1 la cui localizzazione nucleare è a sua volta regolata dall'elevata concentrazione di glucosio. Mig1, che appartiene alla classe degli *zinc finger* C2H2, reprime la trascrizione legandosi ai promotori dei geni che subiscono repressione da glucosio e reclutando altri co-repressori. Se il livello di glucosio è basso Mig1 viene fosforilato dalla chinasi Snf1 proprio nel residuo responsabile del legame con l'esochinasi ed entrambe le proteine rimangono nel citoplasma. In presenza di alti livelli di glucosio l'esochinasi2 interagisce con Mig1 impedendone la fosforilazione e si verifica la traslocazione nel nucleo dove si attiva la repressione genica. Si pensa che l'esochinasi sia anche coinvolta nella regolazione della chinasi Snf1 che regola a sua volta repressori e attivatori della trascrizione (Kim and Dang, 2005) (Gancedo and Flores, 2008).
La lattato deidrogenasi è un tetramero che catalizza la conversione del piruvato in lattato ma è anche in grado di legare il DNA a singolo filamento. Si pensa che il meccanismo alla base del cambio di funzione sia legato alla quantità di substrato dell'enzima poiché il legame al DNA viene inibito in presenza di NADH. Il NADH legandosi all'enzima come co-substrato potrebbe impedire il legame al DNA causando una modificazione conformazionale del tetramero oppure per sovrapposizione dei domini di legami alle due molecole. Il suo ruolo nel nucleo non è ancora ben definito ma si pensa che possa avere un ruolo nella regolazione genica.

La piruvato chinasi è un esempio di proteina che cambia la sua funzione in base allo stato oligomerico: se è presente come omotetramero prende parte all'ultima reazione della glicolisi in cui il fosfoenolpiruvato è convertito in piruvato, mentre sotto forma di monomero agisce come proteina legante ormoni tiroidei e regola la risposta trascrizionale dei recettori di questi ormoni (Sriram et al., 2005).

La gliceraldeide 3-fosfato deidrogenasi è un enzima della glicolisi che converte la gliceraldeide 3-fosfato in bifosfoglicerato ma possiede anche un ruolo nella regolazione della trascrizione in quanto fa parte di un complesso proteico chiamato OCA-S. Questo complesso regola la trascrizione del gene che codifica per l'istone H2B durante la fase S del ciclo cellulare e in particolare la GAPD interagisce con il fattore di trascrizione Oct1 e veicola il complesso sui promotori regolati da questa proteina (Zheng et al., 2003).

Quando si parla di *moonlighting protein* va sottolineato come sia difficile la loro caratterizzazione. Sino ad ora la loro scoperta si può dire essere stata del tutto casuale poiché non esistono delle analisi mirate per l'identificazione di queste funzioni ed è anche difficile poter pensare di sviluppare dei metodi efficaci e attendibili per tale scopo. Inoltre è spesso difficile poter dedurre la seconda funzione di una proteina solamente attraverso una analisi bioinformatica e si è notato che non in tutti i casi si può estendere la nuova funzione alla proteina omologa in altri organismi (Gancedo and Flores, 2008).

Anche se all'interno della cellula le interazioni DNA-proteina sono fondamentali per la regolazione delle funzioni e della sopravvivenza cellulare ci sono ancora molti punti da chiarire per quanto riguarda la comprensione della specifica sequenza di DNA che viene riconosciuta e legata da un fattore di trascrizione e a maggior ragione rimangono ancora tanti interrogativi su quali proteine possano risultare *moonlighting* coinvolte nel legame al DNA (Hu et al., 2009).

Negli ultimi anni sono stati sviluppati dei metodi basati su approcci proteomici per poter studiare proteine in grado di legare il DNA. Nello studio di Hall e collaboratori microarray proteici rappresentanti la quasi totalità delle proteine di *S. cerevisiae* vengono ibridati con il DNA genomico del lievito e in tal modo vengono isolate diverse proteine leganti il DNA tra cui anche proteine non convenzionali. Per poter analizzare in dettaglio queste proteine è necessario condurre esperimenti di immunoprecipitazione. In questo studio è emersa l'inattesa attività di legame al DNA dell'enzima metabolico Arg5,6 che nel mitocondrio è coinvolto nella biosintesi dell'arginina. Arg 5,6 è in grado di legare DNA sia nucleare che mitocondriale e si

pensa che possa avere un ruolo nella regolazione trascrizionale ma anche in quella post-trascrizionale poiché si lega a sequenze presenti negli introni di alcuni geni (Hall et al., 2009).

Nello studio condotto da Hu si utilizzano dei microarray contenenti sia proteine note legare il DNA che proteine con altre funzioni che vengono ibridati con specifiche sonde di DNA marcato risultanti da una analisi *in silico* di sequenze regolatrici. Da questo studio risulta che il 22% delle proteine presenti nell'array per cui non era nota una funzione nella trascrizione sono in realtà in grado di legare il DNA. In particolare si focalizzano sulla chinasi ERK2 per cui era già nota la sua attività chinasica nel nucleo ma per la quale si ipotizza anche un ruolo nella repressione della trascrizione dei geni indotti dall'interferone. ERK2 compete per il sito di legame al DNA con il fattore di trascrizione C/EBP2 che è appunto indotto dall'interferone e quindi possiede un ruolo come regolatore della trascrizione per geni coinvolti nella risposta a stress biotici (Hu et al., 2009).

Tutti questi esempi avvalorano la tesi che le proteine non convenzionali identificate mediante lo *screening* TAT possano non essere tutti dei falsi positivi metodologici ma che in parte possano rappresentare dei reali attivatori della trascrizione. Questa metodologia consentirebbe non solo la validazione degli attivatori trascrizionali ma anche l'identificazione di nuove *moonlighting protein*.

5. I motivi di legame dei fattori di trascrizione

Al crescente aumento della disponibilità delle sequenze genomiche di diversi organismi si sta associando un largo interesse per la comprensione della regolazione dell'informazione genica.

La regolazione della trascrizione è basata su interazioni DNA-proteina ed è per questo che oltre alla classificazione computazionale e funzionale dei fattori di trascrizione diventa fondamentale lo studio e la caratterizzazione delle sequenze regolatrici presenti sul genoma riconosciute da queste proteine che sono definite TFBS (*transcription factor binding site*).

Il promotore di un gene è definito come sequenza regolatrice di poche centinaia di basi posta a monte del sito di inizio della trascrizione. I promotori dei geni che codificano per proteine o per microRNA sono composti da due regioni: un promotore basale e uno prossimale. Il promotore basale serve come sito di riconoscimento del complesso di pre-inizio della trascrizione che posiziona la RNA polimerasi II in prossimità del sito di

inizio della trascrizione mentre il promotore prossimale, localizzato alcune centinaia di basi a monte dal basale, è responsabile della regolazione dell'espressione genica perché riconosciuto dagli specifici fattori di trascrizione. Gli elementi di regolazione in cis riconosciuti da un TF generalmente sono costituiti da corte sequenze degenerate che solitamente non superano le 20bp e che hanno una sequenza consenso in cui determinate posizioni sono fisse mentre altre variabili (Walhout, 2006).

I promotori eucariotici in molti casi si individuano vicino a regioni del genoma chiamate isole CpG che sono corte sequenze ricche in G+C. Sono presenti anche altre zone regolative che a differenza dei promotori sono molto distanti dal sito di inizio della trascrizione e sono rappresentate da enhancer, silencer e isolatori. Gli enhancer sono responsabili della regolazione spazio/temporale della trascrizione genica e la loro funzione è indipendente dalla distanza o dall'orientamento dal promotore. Questi elementi sono generalmente modulari in modo che un promotore possa essere regolato da diversi enhancer in momenti o tessuti diversi o in risposta a diversi stimoli. Gli enhancer inoltre sono composti da un gruppo di siti di legame per i fattori di trascrizione che agiscono in modo cooperativo per promuovere la trascrizione. I silencer hanno le stesse caratteristiche degli enhancer ma sono regolatori negativi della trascrizione dato che interagiscono con i repressori.

Gli isolatori invece sono gli elementi regolativi che agiscono impedendo che un gene possa essere influenzato dalla trascrizione dei geni vicini andando a limitare l'azione dei fattori di trascrizione. Gli isolatori vengono legati da proteine che fungono da "barriera" fisica che impedisce l'azione di un attivatore. Le regioni di controllo del locus (LCR) che vengono riconosciute da fattori di trascrizione, co-attivatori, repressori e modificatori della cromatina sono l'insieme degli elementi regolativi coinvolti nella regolazione dell'espressione differenziale di gruppi di geni. Nella maggior parte dei casi queste sequenze si trovano a monte dei geni che regolano anche se in alcuni casi sono state riscontrate anche all'interno di introni dei geni regolati (Maston et al., 2006).

L'identificazione dei promotori in organismi eucarioti unicellulari come il lievito è meno complessa poiché il genoma è compatto, contiene pochi introni e corte regioni inter-geniche. Lo studio diventa molto più difficile quando ci si approccia a genomi complessi ricchi di sequenze inter-geniche caratterizzate anche da zone composte da sequenze ripetute e in cui sono presenti complessi moduli di motivi di legame al DNA. Per poter decifrare il *network* di regolazione di un fattore di trascrizione non è solo importante capire la sequenza del promotore a cui si lega ma anche identificare la posizione dei questi elementi sul genoma. Un metodo utilizzato per mappare queste regioni è basato

sull'ipotesi che geni co-espressi in determinate condizioni siano regolati dai medesimi fattori di trascrizione. Analizzando *cluster* di geni co-regolati attraverso l'uso di programmi bioinformatici è anche possibile predire possibili motivi di legame comuni alle regioni regolatrici di questi geni. Un secondo metodo di *footprinting* filogenetico si basa sulla conservazione funzionale di elementi regolativi in cis tra organismi strettamente correlati tra loro. Sino ad oggi sono poche le informazioni riguardanti le specifiche sequenze a cui si lega un fattore di trascrizione per questo vi è sempre più crescente interesse per questo argomento. Gli approcci perseguibili sono di tipo computazionale, dove si prevede l'analisi delle regioni regolative di geni co-espressi tramite programmi bioinformatici, o di tipo sperimentale, dove il punto di partenza può essere un fattore di trascrizione che viene analizzato per identificare la sequenza a cui si lega oppure può essere centrato su una sequenza regolativa di interesse che viene analizzata per isolare le proteine in grado di legarla (Walhout, 2006).

A fronte di questo interesse sono state sviluppate numerose tecniche per l'identificazione dei motivi regolativi che consentono l'identificazione del *network* di regolazione di un TF e la scoperta del ruolo di TF a funzione sconosciuta oppure di una proteina *moonlighting*.

Le tecniche si dividono tra quelle che utilizzano approcci in vitro o in vivo. La SELEX è una tecnica in vitro dove uno specifico TF viene incubato con un *pool* di DNA a sequenza casuale e dopo diversi cicli di legame e di lavaggio delle sequenze nucleotidiche non legate si arriva ad un arricchimento delle sequenze specificatamente riconosciute dal fattore di trascrizione (Hudson and Snyder, 2006).

Una seconda metodologia in vitro è la DIP-chip dove un fattore di trascrizione purificato viene incubato con frammenti di DNA genomico e la forma legata viene isolata mediante immunoprecipitazione oppure attraverso purificazione per affinità. Il DNA isolato è amplificato, marcato e utilizzato per ibridare un microarray. Gli svantaggi sono la richiesta di purificazione della proteina e la necessità di utilizzare anticorpi (Liu et al., 2005).

Un altro metodo *high-throughput* in vitro è rappresentato dal *protein binding microarray* e prevede l'utilizzo di un unico array in cui sono rappresentate tutte le possibili sequenze a DNA di una determinata lunghezza. In tal modo l'array non rappresenta più sequenze di uno specifico genoma ma tutte le possibili sequenze di una certa dimensione e può essere applicato a tutti gli organismi. L'array viene ibridato con la proteina di cui si vuole conoscere la specificità di legame a cui è stato fuso un tag che ne permette la visualizzazione mediante l'uso di un anticorpo marcato in modo

fluorescente. Le sequenze che risultano legate vengono poi analizzate per l'identificazione di un motivo di legame conservato (Berger et al., 2006).
Tra le tecniche in vivo una delle più utilizzate è la ChIP-chip in cui una proteina di interesse viene legata covalentemente alla cromatina e successivamente viene immunoprecipitato il complesso proteina-DNA utilizzando anticorpi specifici per la proteina oppure specifici per un tag che viene aggiunto alla sequenza del TF. Il DNA isolato viene amplificato e marcato con sonde fluorescenti in modo da poter ibridare un microarray. La maggiore limitazione del metodo è dovuta dalla disponibilità di un anticorpo specifico per il TF o il dover aggiungere un tag alla sequenza proteica. Inoltre questo metodo anche se permette di evidenziare la sequenza genomica a cui si lega il TF non rappresenta un buon metodo per la misura dell'affinità della proteina a diverse sequenze (Hudson and Snyder, 2006).

5.1 Bacterial one-hybrid

Nel nostro caso abbiamo deciso di intraprendere lo studio delle sequenze regolatrici utilizzando l'approccio del *bacterial one-hybrid* che consente di eseguire l'esperimento in vivo sfruttando un sistema eterologo batterico.
Il sistema è molto semplice, rapido e accessibile a tutti i laboratori poiché è basato su un protocollo di biologia molecolare ampiamente utilizzato. È una tecnica simile a quella condotta in lievito ma ha il vantaggio della maggiore efficienza di trasformazione delle cellule batteriche. Il sistema è formato da tre componenti: un vettore di espressione del fattore di trascrizione (esca), un vettore per la costruzione della *library* composta da una regione a sequenza nucleotidica casuale e un appropriato ceppo batterico per la selezione (Meng et al., 2005).
La *library* viene costruita sintetizzando un frammento di lunghezza definita (dai 18 ai 28 nucleotidi) con sequenza casuale che viene clonato in un vettore dove risulta essere posto a monte di una cassetta che contiene due geni *reporter* posti sotto il controllo di un promotore debole. Quando il TF in esame si lega ad una sequenza della *library* viene reclutata la RNA polimerasi e si attiva l'espressione dei geni *reporter URA3* e *HIS3*. Il gene *URA3* è utilizzato per la selezione negativa che consente l'eliminazione delle sequenze auto-attivatrici cioè di quelle proteine batteriche in grado di attivare l'espressione dei geni *reporter* indipendentemente dall'esca. La selezione è fatta su terreno selettivo addizionato di 5-FOA (*5-fluoro-orotic acid*) che viene convertito in un composto tossico per la cellula dalla orotidina-5-fosfato decarbossilasi codificata da *URA3*. Il gene *HIS3* è coinvolto nella selezione positiva delle sequenze che attivano la

trascrizione in modo dipendente dall'esca e viene effettuata su terreno selettivo privo di istidina e addizionato di 3AT (3-aminotriazolo).

Il vettore dove viene clonato il fattore di trascrizione permette l'espressione in modo inducibile da IPTG di una proteina di fusione tra il fattore di trascrizione e la subunità α oppure ω della RNA polimerasi batterica. È consigliabile, dove possibile, clonare solo il DBD in modo da cercare di evitare eventuali problemi di espressione di una proteina di grandi dimensioni in un contesto eterologo batterico.

Il sistema in cui il DBD è fuso alla subunità α è adatto solo allo studio di alcune classi di DBD come i C2H2 ma non ha portato buoni risultati per i bHLH e gli *homeodomain*. Hanno ipotizzato che questo sia dovuto a una bassa sensibilità del metodo poiché la subunità α è codificata da un gene essenziale e quindi i TF a cui è fusa sono in competizione con le subunità endogene per la formazione del complesso della RNA polimerasi. La subunità omega invece, non essendo codificata da un gene essenziale, permette l'espressione della proteina di fusione TF-ω in un ceppo in cui è stato deleto il gene per tale subunità.

Utilizzando la subunità α è però possibile identificare il sito di legame di proteine che agiscono come dimeri dato che l'RNA polimerasi è composta da due di queste subunità. È infatti disponibile anche un secondo vettore di espressione che garantisce la possibilità dello studio degli eterodimeri.

L'espressione del fattore di trascrizione è inoltre modulabile sia variando le concentrazioni di IPTG nel terreno sia avendo a disposizione tre vettori con promotore debole, medio o forte. Questo permette di risolvere problemi dovuti alla tossicità delle proteine, se troppo espresse, o di ottenere risultati per proteine che danno una forte attivazione dei geni *reporter* solo se altamente espresse nella cellula. La possibilità di utilizzo di diversi promotori garantisce una flessibilità per la caratterizzazione di una vasta varietà di domini di legame al DNA che possono differire per quanto concerne l'affinità di legame, la specificità e i livelli di espressione.

Il sistema di selezione è anche sensibile alla forza di legame del TF, infatti si è notata una correlazione tra dimensione e numero delle colonie su terreno selettivo e l'affinità di legame per quella specifica sequenza. Ci si aspetta che la distribuzione delle sequenze vari in funzione dell'affinità di legame per i diversi TFBS. Le sequenze isolate varieranno anche in base alla stringenza del 3AT, maggiore è la concentrazione maggiore è la selezione di interattori forti, senza sottovalutare che anche le interazioni deboli possono essere biologicamente significative.

Una volta ottenuta la sequenza consenso riconosciuta dal TF se sono disponibili dati di espressione si può determinare se il TF agisce da attivatore o repressore, nel primo caso si troveranno i motivi over-rappresentati in geni che vengono regolati nel medesimo modo rispetto al TF, nel secondo caso i motivi saranno over-rappresentati in geni che hanno un profilo di espressione opposto rispetto al TF (Noyes et al., 2008b).
Questa metodologia è stata ulteriormente modificata per l'analisi di *homeodomain.* Questa classe di fattori di trascrizione non è analizzabile con il metodo classico poiché la loro affinità di legame non è sufficiente a consentire il riconoscimento delle sequenze della *library*. Per poter fronteggiare questo problema hanno costruito dei nuovi vettori per l'espressione dell'esca e della *library* in grado di rafforzare il legame.
L'*homeodomain* viene espresso fuso alla subunità omega della RNA polimerasi e a due domini *zinc finger* (Zif12) del fattore di trascrizione Zif268. La *library* ha una sequenza casuale più corta (10 nucleotidi) poiché è posta immediatamente a monte di un sito di legame noto che viene riconosciuto dal DBD Zif12 che funge da ancora e facilita l'interazione dell'*homodomain*. Con questa strategia si aumenta l'affinità di legame della proteina di fusione *zinc finger–homeodomain* per cui risulta possibile caratterizzare anche i motivi di quei TF che hanno bassa affinità senza creare problemi di *back-ground* poiché il legame del solo Zif12 non è sufficiente per l'espressione dei geni *reporter* (Noyes et al., 2008a).
I principali vantaggi di questa tecnica sono:

- la proteina non deve essere purificata
- l'esperimento è condotto in vivo e sfrutta protocolli rapidi e accessibili e una alta efficienza di trasformazione
- permette lo studio di dimeri
- permette l'eliminazione di falsi positivi.

Sono stati riscontrati anche degli svantaggi ad esempio la complessità che può raggiungere la *library* è di circa 10^8 cloni indipendenti ma non è sufficiente per testare tutte le possibili combinazioni ad esempio di una sequenza di 18 nucleotidi. In base alla lunghezza della *library* bisogna considerare che si possono perdere informazioni riguardo i TF che riconoscono lunghe sequenze. La rimozione delle sequenze auto-attivatrici è indispensabile ma può causare la perdita di sequenze che vengono riconosciute da un fattore di trascrizione di interesse. Ci sono casi in cui è richiesta una seconda ripulitura della *library* per l'eliminazione di tutti i falsi positivi. Inoltre è necessario bilanciare la stringenza del saggio poiché aumentandola si possono perdere tutte quelle sequenze che hanno minore affinità per il TF ma che hanno

ugualmente una rilevanza biologica. Per avere una buona predizione del motivo si dovrebbe procedere al sequenziamento di numerosi cloni anche per eliminare l'influenza di eventuali falsi positivi, l'ideale sarebbe che il metodo fosse seguito da un passaggio di sequenziamento *high-throughtput*. Un altro problema è rappresentato dall'espressione della proteina in cellule batteriche, potrebbero non subire modificazioni post-traduzionali importanti per l'attività del fattore di trascrizione, la proteina potrebbe essere tossica per i batteri oppure potrebbe essere indispensabile modificare codoni della sequenza del DBD se risultassero codoni rari nel batterio. Un ulteriore svantaggio è rappresentato dal clonaggio del solo DBD, il dominio potrebbe non avere la corretta conformazione che normalmente può dipendere anche da altre regioni della proteina (Bulyk, 2005).

Scopo della ricerca

Il mio lavoro di dottorato rientra in un progetto di ricerca internazionale volto alla decifrazione dei genomi di funghi ectomicorrizici, che rappresentano organismi non modello di grande interesse biotecnologico e ambientale.

Dal momento che questi organismi sono difficilmente manipolabili attraverso approcci sperimentali, una parte considerevole del progetto è volta alla messa a punto ed ottimizzazione di sistemi che consentano lo studio della funzione dei geni attraverso metodiche eterologhe.

Lo scopo principale della ricerca è stato la messa a punto di metodiche per lo studio genomico-funzionale dei fattori di trascrizione. Queste proteine, infatti, rivestono un ruolo primario nella regolazione dell'espressione genica, e sono dunque i principali responsabili dell'attivazione dei meccanismi molecolari che stanno alla base dei processi di instaurazione della simbiosi e della formazione del corpo fruttifero.

Lo scopo della ricerca è stato perseguito mediante lo sviluppo di diversi approcci metodologici che comprendono analisi *in silico* e sistemi di *screening* funzionali *in vivo.*

In particolare i principali obiettivi della ricerca sono stati: 1- messa a punto di una *pipeline* per l'identificazione mediante approcci informatici di tutti i potenziali fattori di trascrizione presenti nel genoma di questi funghi; 2- *screening* funzionale nel lievito *Saccharomyces cerevisiae* per l'isolamento di fattori di trascrizione, comprendenti attivatori della trascrizione non convenzionali perché privi di domini di legame al DNA riconoscibili *in silico*; 3- saggio funzionale in lievito per l'identificazione di segnali di localizzazione nucleare criptici; 4- *screening* funzionale in *Escherichia coli* per la determinazione della sequenza di legame al DNA di un fattore di trascrizione di interesse; 5- utilizzo delle metodiche sviluppate negli obiettivi 1-2 per lo studio su scala genomica dei fattori di trascrizione dei funghi ectomicorrizici *T. melanosporum* e *L.bicolor;* 6- utilizzo delle metodiche sviluppate negli obiettivi 3-4 per lo studio approfondito dei fattori di trascrizione e degli attivatori non convenzionali di maggiore interesse perché altamente espressi in una delle fasi del ciclo vitale.

In prospettiva, visto il costante incremento di progetti genomici eucariotici, questo approccio genomico-funzionale potrà essere utilizzato come strumento post-genomico per lo studio dei fattori di trascrizione di altri genomi, sia fungini che appartenenti ad altri organismi.

Parte del lavoro riguardante l'analisi dei fattori di trascrizione codificati dal genoma di *T. melanosporum* è stata pubblicata nell'articolo "*Genome-wide search and functional identification of transcription factors in the mycorrhizal fungus Tuber melanosporum*" B.Montanini, E. Levati et al; New Phytologist (2010).

Materiali e metodi

Ceppi di lievito

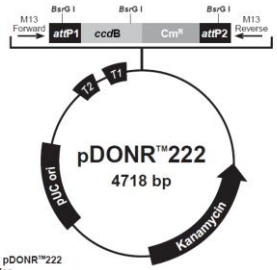
- MaV103 (*MATa, leu2-3,112, trp1-901, his3Δ200, ade2-101, gal4Δ, gal80Δ, SPAL10::URA3, GAL1::lacZ, HIS3UAS GAL1::HIS3@LYS2, can1R, cyh2R*), usato per lo *screening* degli attivatori trascrizionali
- L40 (*MAT a, trp1, leu2, his3, LYS2::lex A-HIS3, URA3::lex A-lacZ*), usato per il saggio di *Nuclear transportation trap*.
- W303-1A (*MATa, leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15)* ceppo utilizzato per la localizzazione mediante fusione a GFP
- CC362-2A (*leu2, ura3, met16*) ceppo utilizzato per le prove di complementazione per la PAPS reduttasi

Ceppi di batteri

- DH10B™ T1R (F- mcrA Δ(*mrr-hsd*RMS-*mcr*BC) φ80lacZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara*,*leu*)7697 *gal*U galK - *rps*L *nup*G *ton*A); usato per la trasformazione e propagazione di costrutti
- DB3.1 (F- gyrA462 endA1 glnV44 Δ(sr1-recA) mcrB mrr hsdS20(r_B^-, m_B^-) ara14 galK2 lacY1 proA2 rpsL20(Sm^r) xyl5 Δleu mtl1) ; usato per la trasformazione e propagazione dei vettori originali Gateway®.
- XL1B (recA1, endA1, gyrA96, thi-1, hsdR17(rK-, mK+), supE44, relA1, lac, [F', proAB, lacIqZΔM15::Tn10(tetr)]) ; usato per la trasformazione mediante elettroporazione dei vettori ottenuti dalla ligazione.
- USO (hisB- pyrF- rpoZ-); usato per lo *screening* del *Bacterial one-hybrid*
- One Shot® TOP10 Electrocomp™ *E. coli* (Invitrogen™) utilizzate per la costruzione della *library* utilizzata per il *Bacterial one-hybrid*

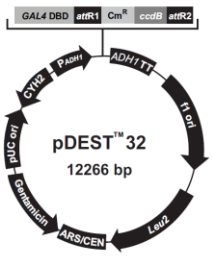
Vettori

- *Transcriptional activator trap* (invitrogen™)



Comments for pDONR™222
4718 nucleotides

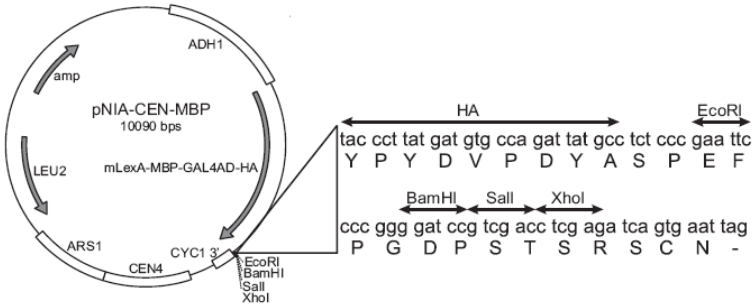
*rrn*B T2 transcription termination sequence: bases 58-85 (c)
rrnB T1 transcription termination sequence: bases 217-260 (c)
M13 Forward (-20) priming site: bases 327-342
*att*P1: bases 360-591
*Bsr*G I restriction sites: bases 442, 1232, 2689
*ccd*B gene: bases 987-1292 (c)
Chloramphenicol resistance gene: bases 1612-2295 (c)
*att*P2: bases 2543-2774 (c)
M13 Reverse priming site: bases 2816-2832
Kanamycin resistance gene: bases 2899-3714 (c)
pUC origin: bases 4045-4718
(c) = complementary strand



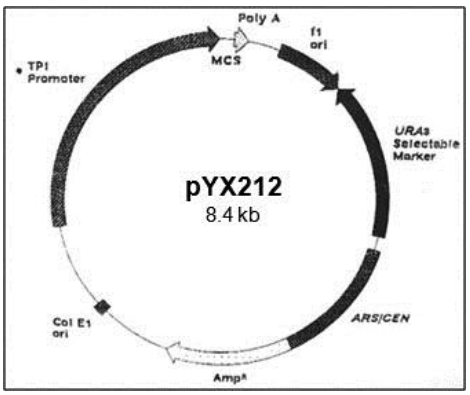
Comments for pDEST™ 32
12266 nucleotides

ADH1 promoter: bases 103-1557
GAL4 DNA binding domain: bases 1581-2024
*att*R1 site: bases 2037-2161
Chloramphenicol resistance (Cm^R) gene: bases 2411-3070
*ccd*B gene: bases 3411-3716
*att*R2 site: bases 3757-3881
ADH1 transcription termination region: bases 4119-4276
f1 origin: bases 4603-5058
Leu2 gene: bases 5767-6861
ARS4/CEN6 origin: bases 7589-8107
Gentamicin resistance gene: bases 8452-8985 (c)
pUC origin: bases 9833-10506
Cycloheximide sensitivity (CYH2): bases 11445-11894 (c)
(c) = complementary strand

- *Nuclear transportation trap* (Marshall et al., 2007)

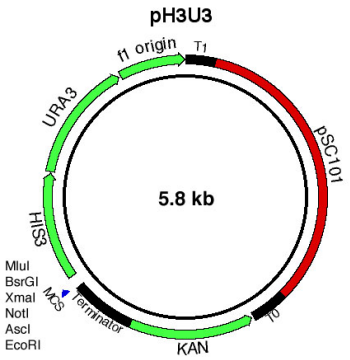


- Analisi di complementazione e localizzione PAPS reduttasi

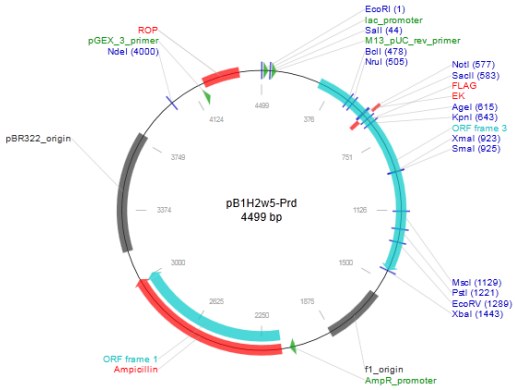


- *Bacterial one-hybrid*

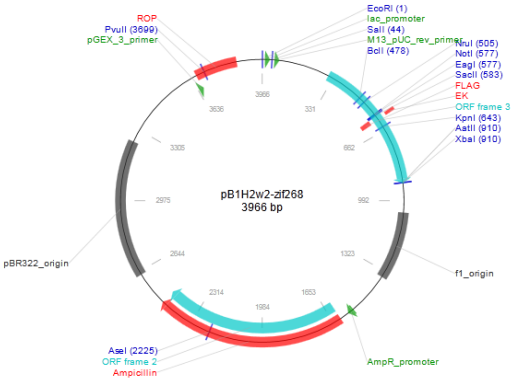
  Vettore per la costruzione della *library* "Addgene plasmid 12609" (Noyes et al., 2008b)



  Vettore per il clonaggio del fattore di trascrizione con promotore medio "Addgene plasmid 18039" (Noyes et al., 2008b)



  Vettore utilizzato come controllo positivo e per il clonaggio del TF con promotore debole "Addgene plasmid 18045" (Noyes et al., 2008b)



Terreni per la crescita dei batteri

Terreno LB**:** triptone 1%, estratto di lievito 5% portato a pH 7.5 con NaOH 2M. Per terreno solido all'occorrenza è aggiunto il 2% di Bacto agar.

Terreno YT2X:

triptone	16gr
estratto di lievito	10gr
NaCl	5gr

Portare a pH 7 con NaOH e portare a 1L con H_2O

Terreno YM: (Meng and Wolfe, 2006)
Autoclavare 9gr di Bacto agar in 451ml di H_2O, quando raggiunge i 55°C circa aggiungere al terreno:

soluzione Sali M9 20X	25ml
estratto di lievito 1%	5ml
glucosio 40%	5ml
uracile 20mM	5ml
istidina 10%	5ml
$CaCl_2$ 0,1M	0,5ml
$MgSO_4$ 1M	0,5ml
Tiamina 1%	0,5ml
$ZnSO_4$ 20mM	0,25ml
5-FOA 12,5mM	2,5ml

Aggiungere al terreno anche gli antibiotici appropriati.

5-FOA: 0,24 gr risospesi in 2,5 ml di DMSO (per 500ml di terreno)

Terreno NM: (Meng and Wolfe, 2006)
Autoclavare 9gr di Bacto Agar in 443 ml di H_2O, quando raggiunge i 55°C aggiungere:

soluzione Sali M9 20X	25ml
glucosio 40%	5ml
adenina 20mM	5ml

mix aminoacidi 33,3X	15ml
uracile 20mM	0,5ml
$CaCl_2$ 0,1M	0,5ml
$MgSO_4$ 1M	0,5ml
Tiamina 1%	0,5ml
$ZnSO_4$ 20mM	0,25ml
5-FOA 12,5mM	2,5ml

Inoltre aggiungere IPTG alla concentrazione finale di 10µM, 3AT in base alla concentrazione desiderata per la selezione e gli antibiotici appropriati.

Miscela aminoacidi 33,3X: preparare 6 soluzioni 200X che poi andranno unite per formare la miscela 33,3X.
Sol I: 0.99 g Phe (0.99%), 1.1 g Lys (1.1%) e 2.5 g Arg (2.5%) in 100ml H_2O
Sol II: 0.2 g Gly (0.2%), 0.7 g Val (0.7%), 0.84 g Ala (0.84%) e 0.41 g Trp (0.41%) in 100ml H_2O
Sol III: 0.71g Thr (0.71%), 8.4 g Ser (8.4%), 4.6 g Pro (4.6%) e 0.96 g Asn (0.96%) in 100ml H_2O
Sol IV: 1.04g Asp (1.04%) e 18.7g potassio-Glu (18.7%) in 100ml H_2O
SolV: 14.6g Gln (14.6%) e 0.36g Tyr (0.36%) in 90ml H_2O. Unire le soluzioni IV e V e aggiungere NaOH per portare in soluzione tutti gli aminoacidi.
Sol VI: 0.79 g Ile (0.79%) and 0.77 g Leu (0.77%) in 100ml H_2O.
Conservare la soluzione a 4°C.

Terreni per la crescita dei lieviti

YPAD 2X: terreno massimo per la crescita dei ceppi di lievito
Bacto yeast extract (2%)
Bacto peptone (4%)
Glucosio (4%)
Adenina solfato (0.004%)
H_2O ad 1 litro

SD 5X

Yeast Nitrogen Base	33.5g
Dropout mix	3.5g

pH 5.6 con KOH

H_2O ad 1 litro

DROPOUT MIX

Adenina solfato	2g
Arginina	2g
Isoleucina	2g
Lisina	2g
Metionina	2g
D-L Serina	4g
D-L Treonina	4g
Fenilalanina	3g
Tirosina	2g
Valina	9g
	32g

Terreno selettivo SD 5X –Leu

Agar	10g
Adenina	20mg
Triptofano	20mg
Istidina	20mg
Uracile	20mg

H_2O fino a 500ml

Autoclavare e aggiungere sterilmente:

Glucosio 20%	50ml
SD 5X	100ml

Terreno selettivo SD 5X –Ura

Agar	10g
Adenina	20mg
Triptofano	20mg
Istidina	20mg

H_2O fino a 500ml

Autoclavare e aggiungere sterilmente:

Glucosio 20%	50ml
SD 5X	100ml

Soluzione SD 5X contenente 3-AT

3-Amino-1,2,4-Triazolo 0,84g

in 10ml di SD 5X per avere una concentrazione di 3-AT 1.0 M

Terreno selettivo SD 5X –Leu+ 3-AT

Agar	10g
Adenina	20mg
Triptofano	20mg
Uracile	20mg
H_2O	350ml

Autoclavare e aggiungere sterilmente:

Glucosio 20%	50ml

e per una piastra dal volume finale di 50ml

[3-AT]	SD 5X	SD 5X / 1 M 3-AT
0mM	10ml	--------
10 mM	9.5ml	500µl
25mM	8.75ml	1.25ml
50mM	7.5ml	2.5ml
100mM	5ml	5ml
200mM	--------	10ml

Transcriptional activator trap (TAT)

Sono stati eseguiti due *screening* per l'identificazione di attivatori trascrizionali utilizzando genoteche preparate a partire da cDNA provenienti da micelio, corpo fruttifero ed ectomicorriza sia del fungo ascomicete simbiotico *Tuber melanosporum* che del basidiomicete simbiotico *Laccaria bicolor* e a partire da cDNA provenienti dalle radici del pioppo *Populus trichocarpa*.

Le genoteche sono state costruite mediante la fusione di cDNA con il dominio di legame al DNA dell'attivatore trascrizionale di lievito GAL4 e sono state inserite nel vettore pDEST32.

La trasformazione delle genoteche è stata effettuata nel ceppo di *S.cerevisiae* MaV103. Questo ceppo contiene tre geni *reporter HIS3, LacZ* e *URA3* che posseggono a monte una regione regolativa *Upstream Activating Sequence* (UAS) di GAL4 riconosciuta dal *DNA binding domain* di GAL4. La trascrizione dei geni *reporter* è attivata solo se il cDNA possiede la funzione di attivatore trascrizionale.

Fig. 1 **:** Rappresentazione della porzione del genoma di MaV103 che contiene i tre geni reporter (immagine da *User Manual* ProQuest™ Two-Hybrid System Invitrogen™)

Il gene *reporter HIS3*, trascritto ad un livello basale, codifica per imidazologlicerol-fosfato deidratasi, il settimo enzima coinvolto nella biosintesi dell'istidina. La sua espressione viene identificata dalla presenza di colonie su terreno di crescita minimo contenente 3-Amino-1,2,4-Triazolo (3-AT), in grado di inibire l'attività di questo enzima.

Il gene *LacZ* codifica per l'enzima β-galattosidasi e la sua espressione è rivelata dalla presenza di colonie blu in presenza del substrato X-Gal, un analogo del galattosio che conferisce la colorazione quando viene metabolizzato dall'enzima.

Il gene *URA3* codifica per l'orotidina-5-fosfato decarbossilasi, un enzima coinvolto nella biosintesi dell'uracile. La sua espressione viene identificata dalla crescita delle colonie su terreno minimo privo della base azotata.

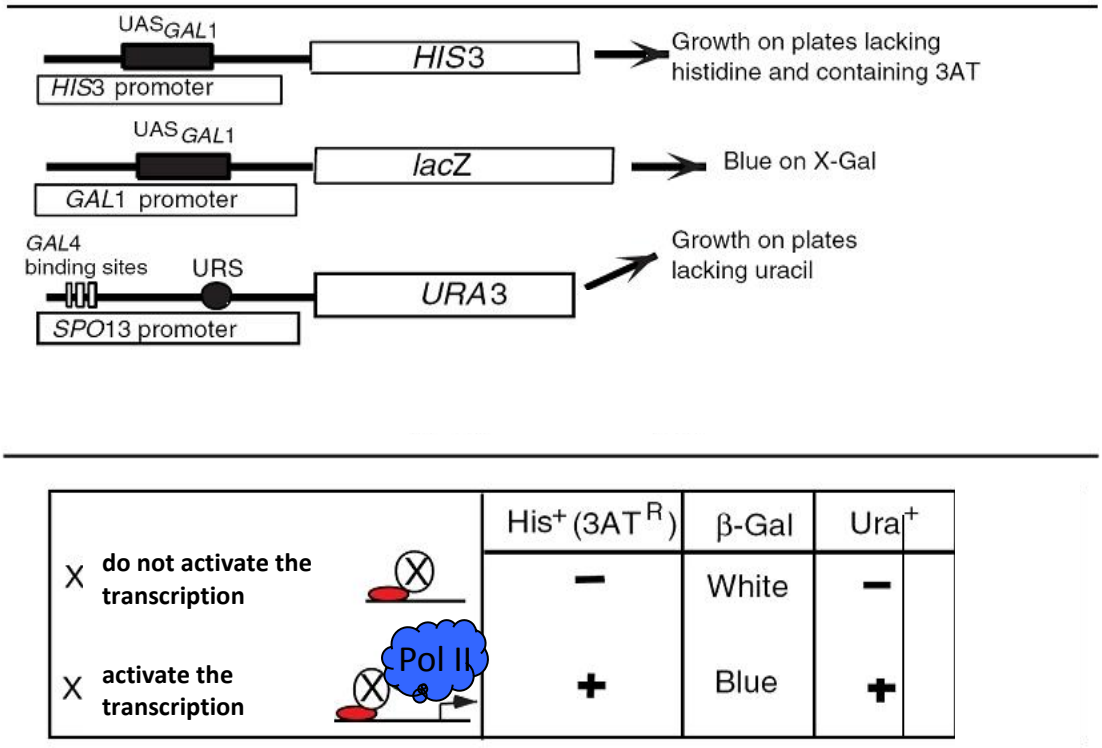


Fig.2: Sistema di identificazione per gli attivatori trascrizionali *reporter* (immagine modificata da *User Manual* ProQuest™ Two-Hybrid System Invitrogen™)

Amplificazione della *library*

L'amplificazione viene fatta in batterio.

- Scongelare la *library* e aggiungerla a 200ml di terreno LB addizionato di gentamicina
- Prelevare 1ml della coltura, diluirlo 1:10 e piastrare 50µl su piastra LB gentamicina per poter poi valutare il titolo iniziale
- Dividere i 200ml in due beute da 100ml ognuna e mettere in agitazione a 37°C per almeno 6 ore dopo le quali si inizia a calcolare l'OD per far raggiungere alla coltura OD compreso tra 0.8 e 1
- Da questa coltura prelevare 25 µl e metterli in 225 µl, mescolare e diluirla ulteriormente 1:10, di questa diluizione piastri su LB gentamicina sia 50 µl che 150µl per calcolare il titolo finale
- Centrifugare la coltura in falcon da 50 ml per 10' a 3000rpm a 4°C e ripetere le centrifugate per tutta la coltura
- Risospendere il pellet in 800µl di TEN e centrifugare per 5' a 4°C, eliminare accuratamente il surnatante e conservare il pellet a -20°C.
- Eseguire midiprep per l'estrazione plasmidica della genoteca

Trasformazione del lievito con DNA plasmidico

La procedura di trasformazione di tutti i ceppi di *S. cerevisiae* utilizzati in questo lavoro di ricerca è la seguente:

- Inoculare una singola colonia del ceppo di lievito che si intende trasformare in 5ml di 2x YPAD e fare crescere in agitazione a 30°C o/n.
- Determinare il titolo della coltura misurando mediante spettrofotometro l'OD a 600nm. La sospensione contenente $8x10^6$ cellule/ml darà un OD pari ad 1.
- Trasferire le cellule in 50ml (per 10 trasformazioni) di 2x YPAD in modo da ottenere un OD iniziale pari a 0,6; incubare in agitazione a 30°C fino ad ottenere un OD=2,4 (4 ore circa).
- Centrifugare le cellule a 3000g per 5 minuti, eseguire 3 lavaggi con acqua sterile.

Aggiungere per ogni aliquota di trasformazione nell'ordine:

salmon sperm DNA	50µl
PEG 3500% w/v	245µl
LiAc 1 M	36µl

Portare a volume finale di 365µl con H_2O. Il *salmon sperm DNA* prima dell'utilizzo deve essere bollito a 100°C per 5'.

- Aliquotare il mix per ogni trasformazione in tubi sterili da 1.5ml
- Aggiungere 5µl di DNA in ogni tubo
- Incubare i tubi a 30°C per 60 minuti
- Incubare i tubi a 42°C per 60 minuti
- Aggiungere 630µl di acqua sterile (volume finale di 1 ml) in ogni tubo e mescolare
- Piastrare 400µl su piastre contenenti terreno selettivo
- Incubare le piastre a 30°C per 3 o 4 giorni e contare il numero dei trasformanti.

Screening su terreni selettivi

Dopo la trasformazione le colonie cresciute maggiormente sono state prelevate dalle piastre di trasformazione e fatte crescere in piastre in formato a 96 pozzetti contenenti terreno SD-Leu per un giorno a 30°C. In ogni piastra sono stati aggiunti anche i controlli wt, m1, m2, pDEST32.

Per valutare l'espressione del gene *HIS3* è stata effettuata la replica delle colonie mediante un *multipinner*, su piastre con terreno selettivo privo di leucina e istidina, addizionato di 3-Amino-1,2,4-Triazolo 25mM e in parallelo sullo stesso terreno ma

addizionati di 3-AT 100mM. Per eseguire la replica liquido-solido, le colonie provenienti dalle piastre in formato a 96 sono state stemperate, sempre grazie al multipinner, in una piastra da 384 o 96 pozzetti contenenti 80µl di H_2O, per poi essere trasferite su terreno selettivo.
Il tempo di incubazione è stato di 2 giorni a 30°C per le colonie cresciute su terreno contenete 3-AT 25mM e 3 giorni per quelle su 3AT 100mM.
L'espressione del gene *HIS3* è stata analizzata catalogando le colonie in base al fenotipo di crescita dividendole in quattro categorie:

- categoria 3: colonie grandi su 3-AT 100mM, indicative di forte espressione del gene *reporter*;
- categoria 2: colonie piccole su 3-AT 100mM;
- categoria 1: colonie che non crescono su 3-AT 100mM ma crescono su 3-AT 25mM;
- categoria 0: colonie che non crescono a causa della mancata espressione del gene *reporter HIS3*.

La valutazione dell'espressione del gene *reporter URA3* è stata effettuata replicando le colonie su terreno selettivo privo di leucina e uracile e facendole crescere a 30°C per tre giorni. Anche in questo caso abbiamo catalogato le colonie in base al loro livello d'espressione, valutando così il fenotipo di crescita:

- 0 per la mancata crescita, quindi mancata espressione di *URA3*
- 1 per colonie di piccole dimensioni
- 2 per colonie di medie dimensioni
- 3 per colonie di grandi dimensioni, quindi forte espressione di *URA3*

Per valutare l'espressione del gene *reporter LacZ*, le piastre sono state replicate su filtro di nitrocellulosa posto su terreno YPAD. Dopo incubazione a 30°C *over night (o/n)*, è stato effettuato il test e si è valutata l'espressione del gene *reporter* in base al grado di colorazione. Sono state catalogate le colonie in 4 livelli di colorazione e quindi di espressione:

- 0 per la colonia bianca (mancata espressione di *LacZ*);
- 1 per la colonia azzurro chiara;
- 2 per la colonia blu chiaro;
- 3 per la colonia blu intenso (espressione massima di *LacZ*).

Per quest'ultimo saggio è necessaria la preparazione della soluzione buffer Z:

$Na_2HPO_4•7H_2O$	16.1g
$NaH_2PO_4•H_2O$	5.5g
KCl	0.75g
$MgSO_4•7H_2O$	0.246g
pH 7.0	

portare a volume di 1 Litro con acqua sterile.

La metodica seguita per il test è stata la seguente:

- Replicare le piastre con i cloni da analizzare su piastre YPAD contenenti un filtro di nitrocellulosa
- Incubare a 30°C o/n
- Per ciascuna piastra, porre un filtro di carta 3M in una piastra petri vuota
- Per ciascuna piastra usare 6ml di buffer Z contenente 11µl di 2-mercaptoetanolo e 100µl di X-Gal al 4%
- Trasferire il buffer Z nella piastra petri contenente il filtro 3M facendo in modo che tutta la carta sia ben bagnata e rimuovendo le bolle d'aria con le pinzette
- Utilizzando le pinzette, prendere i filtri di nitrocellulosa che presentano le colonie di lievito cresciute e congelare i filtri in azoto liquido
- Scongelare a temperatura ambiente e appoggiare i filtri sulla carta 3M rimuovendo le bolle d'aria (tenere le colonie cresciute verso l'alto)
- Incubare a 37°C o/n

Colony PCR su lievito

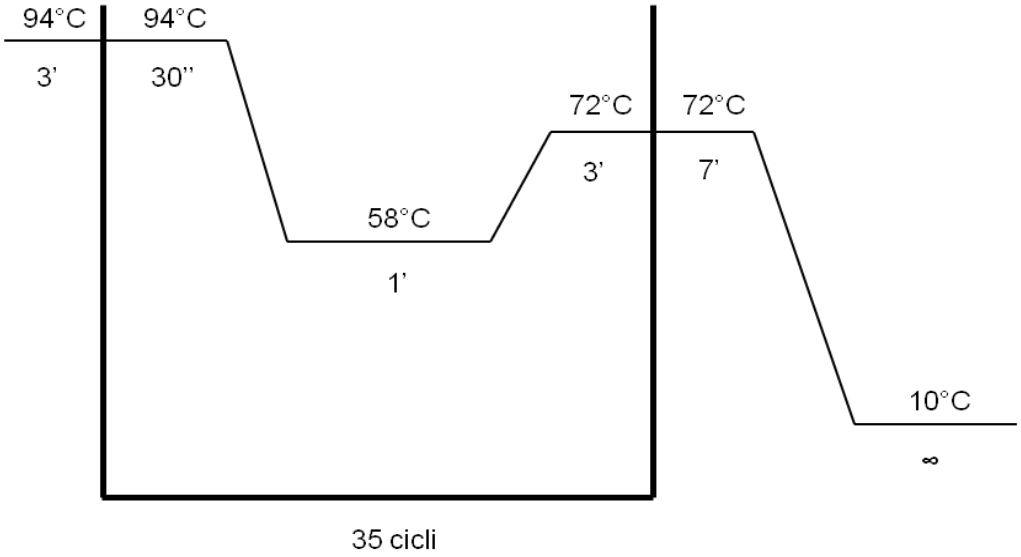
Questa metodologia è stata ottimizzata durante il mio lavoro di tesi per identificare le migliori condizioni di reazione. Una volta cresciuti i lieviti, preparare per ogni colonia un tubo da 0.2ml contenente 10µl di NaOH 20mM preparato fresco da uno stock 2 M. Prelevare le singole colonie di lievito e risospenderle in NaOH 20mM. In questo caso è stata valutata la migliore densità del lievito per ottenere una buona amplificazione andando a monitorare l'efficienza di reazione cambiando il valore dell'OD delle cellule inoculate. Abbiamo considerato come massimo un OD= 12 e poi effettuate diluizioni 1:3, 1:10, 1:30, 1:100, 1:300 e 1:1000. Da questo test si conclude che una alta concentrazione di lievito non inibisce la reazione mentre ad una bassa concentrazione non otteniamo amplificazione. Dopo aver scaldato i campioni per 15 minuti a 98°C in un termociclatore si deve aggiungere ad ogni campione 10µl di Tris-HCl. In questo

caso sono state testate diverse concentrazioni di Tris-HCl: 0.25, 0.5, 1 e 2M. Dai risultati ottenuti si è potuto concludere che le concentrazioni migliori sono 0.25 e 0.5 M e che aumentando la concentrazione di Tris-HCl diminuisce l'efficienza di amplificazione. È poi stata testata la migliore concentrazione di primer da utilizzare nella reazione utilizzando 1µl oppure 1,25µl di primer 10mM e si è dedotto che la corretta quantità è 1µl. Sono state fatte anche delle prove modificando la quantità di DNA aggiunta al mix di reazione, abbiamo ripetuto la reazione con 1.5, 2, 2.5, 3, 3.5, e 4 µl di DNA in un volume dimezzato rispetto al volume di reazione per una PCR preparativa. È risultato che la quantità migliore è 3 µl in 25 µl finali di conseguenza in un volume finale di 50 µl è meglio aggiungere 6 µl di DNA. Preparazione del mix di reazione della PCR, considerando come volume di DNA 6µl in 50µl finale

Protocollo modificato:
Preparare per ogni colonia di lievito da analizzare un tubo da 0.2ml contenente 10µl di NaOH 20mM preparato fresco da uno stock 2M.
Prelevare le singole colonie di lievito e risospenderle in NaOH 20mM.
Scaldare per 15 minuti a 98°C in un termociclatore.
Trascorsi i 15 minuti aggiungere ad ogni campione 10µl di Tris-HCl 0.25M pH8, mescolare bene e poi aliquotare 6µl nel mix di reazione:

MIX per Colony PCR	1 colonia
Buffer 5X	10 µl
dNTPs 10mM	1 µl
primer Fw 10µM (BM99)	1 µl
primer Rev 10µM (BM188)	1 µl
DMSO	2.5 µl
Go-Taq® (5U/µl)	0.25 µl
H_2O	28.25 µl
Volume finale MIX	44 µl

Procedere con i seguenti cicli di PCR:



Caricare su gel di agarosio 3µl di prodotto di PCR per verificare la presenza degli ampliconi, compreso il controllo negativo. I prodotti di PCR sono stati purificati mediante il sistema Millipore® prima di essere sequenziati.

Miniprep di lievito con zimoliasi

- inoculare la colonia di lievito in 10ml di terreno liquido YPAD 2X e crescita in agitazione a 30°C per over night
- centrifugare 3ml di coltura per 5 minuti a temperatura ambiente e rimuovere il surnatante
- risospendere in 1ml di TEN
- centrifugare per 5 minuti e rimuovere il surnatante
- risospendere in 240µL di soluzione I contenente RNasi 10 µg/mL
- aggiungere 10µl di zimoliasi (0,5u/µL) e 5µl di 2-mercaptoetanolo e risospendere
- incubare a 37°C per 30 min
- aggiungere 240µl di soluzione II (0.2N NaOH+ 1% SDS), che va preparata fresca, e mescolare per inversione senza scuotere per 4-8 volte
- incubare per 3-5 minuti
- aggiungere 300µl di soluzione III e mescolare per inversione per 4-8 volte
- centrifugare a 4°C per 10 minuti e prelevare il surnatante

- aggiungere al surnatante 0,6 volumi di isoprapanolo e centrifugare per 10 minuti a temperatura ambiente
- eliminare il surnatante ed aggiungere 1ml di etanolo 70%
- centrifugare per 10 minuti ed eliminare tutto l'etanolo
- risospendere in 4µl di acqua.

Infine viene effettuata una PCR su miniprep di lievito per ottenere l'amplicone da mandare a sequenziare.

Primer utilizzati

Nome	Codice	**Sequenze**	F / R	Tm	Note
2H-Re	BM186	ACCAACGATTTGACCCTTTTCCA	R	66	Primer utilizzato per sequenziamento Rev
2H-BD-Fw	BM99	CAGTGGAGACTGATATGCCTCTAAC	F	60	Primer utilizzato per sequenziamento Fw

Analisi bioinformatiche

Ottenuti i risultati del sequenziamento, le sequenze sono state analizzate tramite metodi bioinformatici. Per prima cosa è stato utilizzato il software MEGA per riconoscere ed eliminare la sequenza di circa 150 nucleotidi appartenente al vettore. Le sequenze risultanti sono state convertite in formato FASTA ed analizzate con il programma EGassembler (http://egassembler.hgc.jp/) per essere assemblate in *contig* ed ottenere delle sequenze uniche. Queste ultime sono state associate ai rispettivi *gene model* di *Laccaria bicolor* e di *Populus trichocarpa* mediante un BlastX (http://www.ncbi.nlm.nih.gov/blast/) mentre le sequenze di *Tuber melanosporum* sono state analizzate attraverso la banca dati EST del sito *Tuber* MycorWeb (http://mycor.nancy.inra.fr/IMGC/*Tuber*Genome/index.html).

Le proteine corrispondenti sono quindi state analizzate per ottenere informazioni sulla loro funzione, mediante due metodi: la ricerca di domini conservati (http://www.ebi.ac.uk/Tools/InterProScan/) e BLASTP per identificare proteine omologhe a funzione conosciuta (http://www.ncbi.nlm.nih.gov/blast/). La presenza di segnali di localizzazione nucleare è stata condotta utilizzando PSORTII (http://psort.ims.u-tokyo.ac.jp/form2.html).

TAT mirato per attivatori trascrizionali codificati dal genoma di *T. melanosporum*

Per validare funzionalmente alcuni tra i TF che sono stati isolati dall'analisi *array* e dall'analisi RNA-seq per essere espressi in maniera preferenziale in corpo fruttifero ed

ectomicorriza rispetto al micelio, le sequenze sono state clonate in pDEST32 in modo da verificare il fenotipo di attivazione dei tre geni *reporter*.
Le amplificazioni dei TF sono state effettuate tramite *touch-down PCR* partendo dalle genetoche in cui erano espressi maggiormente. I *primer* utilizzati contenevano siti di restrizione CpoI, in modo da avere clonaggi direzionali.

Name	espr	Nome	length	primer	sequenza	Tm
TmelACUK	FB	TF1	1950	BM152	caCGGTCCGcagtatttcatcATGGTGAAGC	57
				BM153	caCGGACCGgtaagaaaacaaaaTTACTTCCCCC	57
HP-DBD	ECM	TF2	2200	BM154	ccCGGTCCGATGCCCAGCGACGACCA	60
				BM155	ccCGGACCGaTTACCGGGAGCCAGAGC	60
TmelCBF1	FB	TF3	1050	BM156	ccCGGTCCGtacaaaccaATGCCGATGG	60
				BM157	ccCGGACCGatgTCACGATGAGCCTCCA	60
CH-TF	ECM	TF4	1300	BM158	ccCGGTCCGATGTCGGACTTTGGAGATCAT	58
				BM159	ccCGGACCGTTATGTACCTTGTGGGGTGTGC	58
HP-DBD	ECM	TF5	1400	BM160	ccCGGTCCGGCAATGCTCTCGCAACCT	60
				BM161	ccCGGACCGCTATGCAAGGCAGAAATCGTCT	60
TmelFF7	ECM	TF6	1800	BM162	ccCGGTCCGACCATGGGTGAAATCAGGAG	60
				BM163	ccCGGACCGTTCACATATCCGTGCTATCATTG	60
CH-TF	FB	TF7	2100	BM184	TACAAAAAAGTTGGCGGTCCGAtggattttggaggagccta	59
				BM185	AAGAAAGTTGGGTCGGACCgacactaaaagtactgcatatcaaagc	59
HP-DBD	FB	TF8	1850	BM166	ccCGGTCCGATGACTCGCCATCCGCT	60
				BM167	ccCGGACCGctCTAGTCCCCCATGTGAGC	60
HP-DBD	ECM	TF9	1245	BM168	ccCGGTCCGATGGATAGCTACTTTCCGTCTAATC	58
				BM169	ccCGGACCGTCACCTCTCCCAAGTATTGTTC	58
TmelXLNR-2	ECM	TF10	1150	BM187	GGGGACAAGTTTGTACAAAAAAGCAGGCTtaATGCACCACACACAAAACTCC	62
				BM188	GGGGACCACTTTGTACAAGAAAGCTGGGTctcccgattgtattcgcact	62
CH-TF	ECM	TF11	3250	BM174	ccCGGTCCGATGGAGTCCCAATTGCCA	60
				BM175	ccCGGACCGctaaacacccATGctcaaattca	60
CH-TF	ECM	TF12	3000	BM176	ccCGGTCCGATGCGGTCGTCTATTGCTTG	61
				BM177	ccCGGACCGCTACTTCTCCCACATCCCATCA	61
HP-DBD	FB	TF13	1150	BM178	ccCGGTCCGATGCCTAACAGGATGGAACC	59
				BM179	ccCGGACCGCTATTTTTGAGTGACGAGTGTGAAC	59
CH-TF	ECM	TF14	2300	BM180	caCGGTCCGATGtttcgcaacaggcagtc	61
				BM181	caCGGACCGgTTAATATATGAGTCTCAGCTCC	61
CH-TF	ECM	TF15	1850	BM182	ccCGGTCCGATGCTCGCAGCGGTAGAC	59
				BM183	ccCGGACCGtctccatcttctccTTAAATAAACAA	59

Tabella 1: primer utilizzati per l'amplificazione dei TF da clonare nel vettore pDEST32

Gli ampliconi sono stati sottoposti a digestione con l'enzima di restrizione CpoI e successivamente ligati nel vettore pDEST32 digerito CpoI ad eccezione del TF7 e TF10 che sono stati clonati tramite il metodo Gateway® perché il clonaggio classico non dava buoni risultati. Mentre non si è ottenuta amplificazione dei fattori di trascrizione TF11 e TF15.

Le reazioni di ligazione sono state utilizzate per trasformare cellule batteriche elettrocompetenti DH10B™ T1R e dopo crescita su terreno LB addizionato di gentamicina, i trasformanti sono stati analizzati mediante *colony* PCR batterica.
Sono state effettuate delle minipreparazione plasmidiche dei clonaggi, sono stati purificati mediante purificati mediante *Kit Wizard® SV Gel and PCR Clean-Up System* (PROMEGA) ed inviati presso la Macrogen per il sequenziamento.
I cloni sono stati utilizzati per la trasformazione in lievito MaV103. Per valutare l'attivazione dei geni *reporter* è stato fatto il *drop assay* partendo da concentrazione iniziale di 1 OD e poi delle diluizioni seriali fino a 10^{-4} OD delle cellule di lievito trasformate. Il saggio è stato condotto su terreno SD privo di leucina ed uracile per valutare l'espressione del gene *URA3* e su terreno SD privo di leucina ed istidina addizionato con 3-AT alle concentrazione di 10, 25, 50, 100 e 200 mM per il gene *HIS3.* Per *LacZ* invece, la concentrazione 0,1 OD è stata depositata su filtro di nitrocellulosa posto su terreno YPD. In parallelo sono stati piastrati i controlli: wt, m1, m2 e il vettore vuoto come controllo negativo (pDEST32/CpoI).
Ogni fattore di trascrizione è stato classificato come attivatore debole, medio e forte oppure non attivatore in base ai fenotipi di crescita sui terreni selettivi. Se è in grado di crescere sino a 100-200 mM 3AT, cresce su SD senza uracile ed è blu nel saggio *LacZ* è classificato come forte, se cresce sino a 50mM e nel saggio *LacZ* si ottiene colorazione si classifica come medio, gli attivatori deboli crescono solo sino a concentrazione 25-50mM di 3AT e non sono positivi a saggio della β-galattosidasi e non crescono in assenza di uracile.

Touch-down PCR

Gli inserti corrispondenti ai TF sono stati amplificati mediante *touch-down* PCR utilizzando come templato genoteche a cDNA del ciclo vitale in cui sono maggiormente espressi.Il programma di PCR è stato impostato in modo che durante i primi 10 cicli la temperatura di *annealing* diminuisca di un grado ad ogni ciclo (da 65°C a 55°C), successivamente vengono eseguiti 35 cicli alla temperatura di 55°C.

Mix di reazione:

cDNA	2 µl
Buffer HF Phusion 5X	10 µl
dNTPs (10 mM)	1 µl
primer Fw (10 µM)	1 µl

primer Rev (10 µM)	1 µl
Taq Phusion®	0.5 µl
H_2O	34.5 µl
Vfinale	50 µl

Reazioni di ricombinazione Gateway®

La tecnologia Gateway® é una metodologia di clonaggio universale che si basa sulle proprietà della ricombinazione sito-specifica del batteriofago Lambda e costituisce una via veloce a altamente efficiente per il trasferimento di sequenze di DNA in vettori di interesse.

Le reazioni di ricombinazione che compongono questa tecnologia sono due:

- la reazione BP che facilita la ricombinazione di un substrato *att*B (un prodotto di PCR-*att*B o un clone di espressione *att*B linearizzato) con un substrato *att*P (vettore donor) per generare un clone *att*L. Questa reazione é catalizzata dall'enzima BP Clonase™ II;

- la reazione LR che facilita la ricombinazione di un substrato *att*L con un substrato *att*R (vettore di destinazione) per creare un clone di espressione *att*B. Questa reazione é catalizzata dall'enzima LR Clonase™ II.

In questo lavoro di tesi si é utilizzato pDONR222 come vettore donor e pNIA come vettore di destinazione per eseguire la Yeast Nuclear Targeted Trap (NTT).

I vettori utilizzati in queste reazioni, quando provenienti da miniprep effettuati manualmente, sono stati purificati nel seguente modo:

- portare ad un volume di 200µl con acqua
- aggiungere 200µl di PEG 13% NaCl e vortexare

- centrifugare 10 minuti a temperatura ambiente
- risospendere il pellet in 200µl di TE
- aggiungere NaAc (1/10 del volume) e due volumi di EtOH 100%, vortexare
- centrifugare 10 minuti a 4°C
- fare un lavaggio con EtOH 70% e risospendere in 30µl di TE

Reazione BP

DNA (pDEST32)	30ng
pDONR222	30ng
BP Clonase™ II (5X)	0.2µl
TE fino ad 1µl	

Incubare a 25°C o/n

- Aggiungere la soluzione proteinasi K (1/10 del volume di reazione: 0.2µl) e incubare a 37°C per 10 minuti
- Trasformare mediante trasformazione chimica con 0.45µl di reazione cellule DH10T1R e piastrare su terreno selettivo contenente kanamicina
- Purificare mediante miniprep su batterio i costrutti in pDONR222

Reazione LR

DNA (pDONR222)	30ng
pNIA	30ng
LR Clonase™ II (5X)	0.4µl
TE fino a 2µl	

Incubare a 25°C o/n

- Aggiungere la soluzione proteinasi K (1/10 del volume di reazione: 0.2µl) e incubare a 37°C per 10 minuti
- Trasformare mediante trasformazione chimica con 0.45µl di reazione cellule DH10T1R e piastrare su terreno selettivo contenente ampicillina. Se il numero di trasformanti non fosse sufficiente, utilizzare l'elettroporazione
- Purificare mediante miniprep su batterio

Successivamente, questo DNA purificato viene utilizzato per trasformare il ceppo di lievito Mav103.

Nuclear transportation Trap (NTT)

Per validare la presenza di un segnale di localizzazione nucleare all'interno dei *"Putative unconventional factors"* e degli "*Unconventional activator*" è stato adottato l'NTT.

Il vettore utilizzato pNIA-CEN-MBP, è un plasmide centromerico che permette l'espressione in lievito di una tripla proteina di fusione che contiene la sequenza batterica di *LexA* modificata (*Δ*NLS), il dominio di attivazione Gal4AD e il cDNA della proteina da testare sub-clonata in frame a valle di Gal4AD. La trasformazione di questi vettori è stata effettuata nel ceppo di lievito L40 di *S. cerevisiae* contenente due geni *reporter*: *HIS3 e LacZ,* i quali posseggono a monte il sito operatore LexA riconosciuto dal *DNA binding domain* di LexA. Di conseguenza se la sequenza presa in considerazione ha un NLS proprio, la proteina di fusione espressa entrerà autonomamente nel nucleo attivando l'espressione dei geni *reporter HIS3 e LacZ*. Come controllo negativo è stato utilizzato il vettore pNIA-CEN-MBP.

Per l'amplificazione delle sequenze sono state utilizzate le genoteche a cDNA derivanti dalle tre fasi del ciclo vitale del fungo. L'amplificazione da genoteca consente di isolare l'intera sequenza perché in alcuni casi i clono isolati dal TAT hanno solo una sequenza parziale del gene. I primer utilizzati contengono al loro interno i siti di restrizione opportuni per il clonaggio. Gli ampliconi dopo eluizione da gel di agarosio sono stati digeriti con l'enzima di restrizione opportuno e clonati mediante ligazione all'interno del vettore pNIA-CEN precedentemente digerito. La reazione di ligazione è stata utilizzata per la trasformazione mediante elettroporazione del ceppo batterico XL1B e i trasformanti sono stati piastrati su terreno LB addizionato di ampicillina. Le colonie cresciute sono state sottoposte a *colony* PCR per identificare i cloni contenenti il costrutto corretto dai quali si è poi estratto il vettore che è stato utilizzato per la trasformazione del lievito L40.

La presenza del segnale di localizzazione nucleare è stata saggiata mediante *Drop assay* su terreno SD privo di leucina e istidina e addizionato di 3AT alle concentrazioni 5, 10, 25, 50mM per verificare l'espressione del gene *reporter* HIS3 e su filtro di nitrocellulosa per effettuare il saggio per la β-galattosidasi.

gene model	classe	Sc homolog	codice	sequenza primer	Tm
GSTUMT00008609001	pTFs	MetR	BM62	cccGGATCCaatggctggaacctacagc	56
			BM41	CGCctgcaGTCGAC TTTGTACAAGAAAGCTGGGT	56
GSTUMT00001950001	UAs	Ltv1	BM131	CGggatccTCATGCCAACCAAAAAGTGGATC	60
			BM132	TCgtcgacTCAAGAAGCGATCCTCGCAC	60
GSTUMT00006162001	UAs	Act1	BM117	CGgaattcATGGAAGACGAAGTTGCCGC	60
			BM118	Tcggatcccacgttaaatcacctcttatac	60
GSTUMT00004791001	UAs	Mxr1	BM291	TAAATATAAAagatctTAATGTTCACATCTCTCCTCCCTC	60
			BM292	AATTATTTTAagatctTCAAGCGAGCGGTGGAA	63
GSTUMT00003999001	PUAs	Lap3	BM293	TAAATATAAAGAATTCATGGGGTCGTCGTCATCTAA	60
			BM294	AATTATTTTActcgagCTAAGCAAGGCTTCCCATTG	60
GSTUMT00004644001	PUAs	Pdh1	BM295	TAAATATAAAagatctTA ATGTCATTCCTTCCTCGCC	60
			BM296	AATTATTTTAagatctTTATCCCTTAGTAACATACAAATCG	56
GSTUMT00008696001	PUAs	Zwf1	BM297	TAAATATAAAggatccTCATGGCAGACACTATTATTCAATCC	59
			BM298	AATTATTTTAgtcgacCTAGAGCTTGTGAGGAGCAGTG	59
GSTUMT00002663001	PUAs	Met16	BM66	cc gaa ttc atg acagtaactccaactga	56
			BM64	CGCc tcg agatct TTGTACAAGAAAGCTGGGT	54
GSTUMT00003049001	PUAs	Mtd1	BM301	TAAATATAAAGAATTCATGTCCACAATAGTAGAGGACCC	58
			BM302	AATTATTTTAGGATCCCTACAGACTCTCCCCTCCCT	58
GSTUMT00002382001	PUAs	Pex20	BM43	cgg atc cag atg gcgctttgcgggcc	58
			BM41	CGCctgcaGTCGAC TTTGTACAAGAAAGCTGGGT	56
GSTUMT00005390001	UAs	Maf1	BM307	TAAATATAAAggatccTCATGAAGTTCCTTCCTCTTCGC	60
			BM308	AATTATTTTAgtcgacTCAAACTTCCATCCCTTCG	59
GSTUMT00010173001	PUAs	Ste50	BM313	TAAATATAAAggatccTCATGCCGTCTCCGGGTCATCA	64
			BM314	AATTATTTTAgtcgacTTATAATACTCCACCAGGTAATGTGA	58
GSTUMT00005170001	PUAs	Cnb1	BM135	CGggatccTCATGGCTCAACAACAGTGGATC	62
			BM136	TCgtcgacTCAATGGTGTACTCCGTTAGT	60
GSTUMT00003548001	PUAs		BM125	CGagatctTAATGGCTTCGTCACTGCGTG	60
			BM126	TCagatctTCAAATCCAATCACCGCCCT	60
GSTUMT00007946001	PUAs		BM129	CGagatctTAATGTCTCCTCAAGTCAAACCC	62
			BM130	TCgtcgacTCAGTCTAGGGGATCGGCT	60
Eugenev2.0000060193	LbTASP1		BM260	CGggatccTCGAGCAACATCATAGTGCTCCA	59
			BM261	TCgtcgacTTAATCCAACTCGTCAACATCA	58
eu2.Lbscf0025g01210	LbTASP3		BM275	ccGCGGTCCGGAATTCCTCCCCATGGGTGCCGA	67
			BM276	ccGCGGACCGGATCCTTAAAGACGAGAGACGGCAGTAG	60

Tabella 2: primer utilizzati per l'amplificazione degli inserti da clonare nel vettore pNIA-CEN-MBP (Sc: *S.cerevisiae)*

Bacterial one-hybrid

Il *bacterial one-hybrid* è un sistema che permette di determinare la specificità di sequenza di un fattore di trascrizione. Questo sistema di selezione prevede l'utilizzo di due vettori uno esprimente il fattore di trascrizione fuso alla subunità ω della RNA polimerasi batterica e uno al cui interno viene clonata la libreria di sequenze casuali. Entrambi i vettori verranno utilizzati per trasformare il ceppo batterico USO che è deleto nei geni *hisB* e *pyrF* omologhi a *HIS3* e *URA3* e nel gene che codifica per la subunità omega della RNA polimerasi. Quando il fattore di trascrizione riconosce e lega una sequenza all'interno della libreria recluta la RNA polimerasi batterica e promuove l'espressione dei geni *reporter*.

Il gene *reporter URA3* viene utilizzato per eseguire la ripulitura della *library* dalle sequenze auto-attivatrici, cioè da tutte quelle sequenze della library che provocano l'attivazione dei geni *reporter* in assenza del fattore di trascrizione. La selezione viene condotta in terreno YM contenente 5-FOA che viene convertito in un composto tossico dal prodotto genico del gene *URA3* di conseguenza in tali condizioni non dovrebbero crescere le colonie auto-attivatrici.

Il gene *reporter HIS3* consente la selezione delle sequenze casuali che vengono legate dal fattore di trascrizione in esame. Quanto il TF si lega alla sequenza provoca l'attivazione dell'espressione genica del gene *HIS3* che conferisce fenotipo di crescita su 3AT.

Il protocollo su cui mi sono basata è descritto da Meng e collaboratori (Meng and Wolfe, 2006).

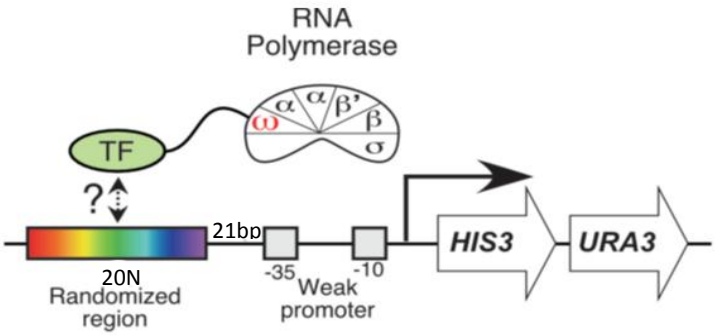


Figura 3 : sistema di selezione del B1H (Noyes et al., 2008b)

Clonaggio del fattore di trascrizione

Il fattore di trascrizione è espresso come fusione carbossi-terminale alla subunità ω della RNA polimerasi batterica ed è sotto il controllo di un promotore che ne permette

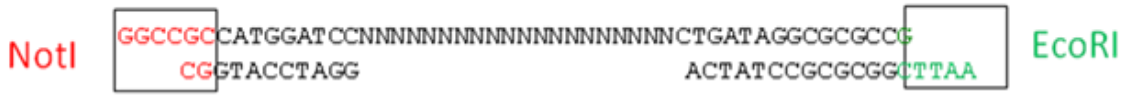
l'espressione inducibile da IPTG. Sono disponibili tre tipologie di vettori per il fattore di trascrizione che si differenziano per il promotore sotto il controllo del quale è posto il TF. Negli esperimenti condotti è stato utilizzato il vettore pB1H2ω5 contenente il promotore *lpp/lacUV5* che consente un livello di espressione moderata e il plasmide pB1H2ω2 con il promotore *lpp/lacUV2* che provoca livello basso di espressione. Quest'ultimo vettore viene generalmente utilizzato quando livelli troppo alti della proteina risultano tossici per la cellula batterica. Esiste anche un terzo tipo di vettore, non utilizzato negli esperimenti, che possiede un promotore per un alto livello di espressione. I fattori di trascrizione amplificati da genoteche di *Tuber* provenienti da micelio, corpo fruttifero o ectomicorriza vengono clonati all'interno del plasmide pB1H2ω5 o del pB1H2ω2 mediante clonaggio con gli enzimi di restrizione KpnI e XbaI. Dove è possibile si è clonato solo il dominio di legame al DNA cercando di lasciare 10 amminoacidi addizionali alle estremità del dominio per essere sicuri di clonare l'intero dominio funzionale. Nel caso in cui fosse nota la omodimerizzazione del TF è opportuno cercare di clonare anche il dominio di dimerizzazione anche se spesso la predizione di tale regione risulta difficile. Una volta ottenuto il costrutto si trasforma il ceppo batterico USO.

gene model	nome	primer	sequenza	Tm	bp
GSTUMT00002283001	NIRA1-DBD	BM212	**cccGGTACCggt**CCGCTGAAAAGACGCTGTGTCAGTAA	60	354
GSTUMT00002283001	NIRA1-DBD	BM213	**cctTCTAGACGGACCG**TTAGCCATCCGGCGATGCCTC	70	354
GSTUMT00002283001	NIRA1	BM214	TACAAAAAAGTTGGC**GGTACCggtCCG**ATGGCAGGCGTTACCTTTG	61	2754
GSTUMT00002283001	NIRA1	BM215	AAGAAAGTTGGG **TCTAGA**CGGACCgTCACTCCTCCCCGTCGTC	62	2754
GSTUMT00003209001	NIRA2-DBD	BM216	TACAAAAAAGTTGGC**GGTACCggt**AAGCGACGCTGCGTGTCTAGT	64	339
GSTUMT00003209001	NIRA2-DBD	BM217	AA GAA AGT TGG G **TCTAGA**TTACGAATGCACCTCACAATCTGG	68	339
GSTUMT00009814001	TF8-DBD	BM218	TAC AAA AAAGTTGGC**GGTACCggt**CAAGTAGCTTGTGGGCGATG	62	405
GSTUMT00009814001	TF8-DBD	BM219	AAGAAAGTTGGG**TCTAGA**TTAGCTACCCAGAATCAGACCAC	57	405
GSTUMT00004776001	TF13-DBD	BM220	**cccGGTACCggt**CCCTACGCTGTTTTGATCTATCG	62	288
GSTUMT00004776001	TF13-DBD	BM221	**cctTCTAGA**TTAAGCATTGGGTGGAGACTTCC	61	288
GSTUMT00005836001	TF5-DBD	BM222	TACAAAAAAGTTGGC**GGTACCggt**GCTGAGGACTGCTTCACTTG	59	402
GSTUMT00005836001	TF5-DBD	BM223	AAGAAAGTTGGG**TCTAGA**TTAAGAATTTGACGGAGGTTCCA	59	402
GSTUMT00012451001	XLNR-2	BM224	**cccGGTACCggt**atgcaccacacacaaaactcc	61	957
GSTUMT00012451001	XLNR-2	BM225	**cctTCTAGAtca**cctatcatccacacaacc	58	957
	Sc-Rds1-DBD	BM264	**CccGGTACCggt**ATGGATTCGATTACAGTAAAAAAAC	57	2499
	Sc-Rds1-DBD	BM265	**CctTCTAGA**TTATGTAACTAATCCTACTATAACTAGCATATC	54	336
	Sc-Met32	BM266	**CcGGTACCggt**CCGATGGAGGATCAGGATGCTG	58	576
	Sc-Met32	BM267	**CtTCTAGACGGACCG**TCAGCCATTACTGCTACCATTG	60	576
GSTUMT00012172001	Met32-FL	BM268	**CccGGTACCggt**ATGAGCACCCACGACATGTC	62	1242
GSTUMT00012172001	Met32-FL	BM269	**cctTCTAGA**TTATCGGTCATGCTTTATCATACCAC	62	1242
GSTUMT00012172001	Met32-DBD	BM270	**CccGGTACCggt**CGACCCTTCAAGTGTGACCA	62	327
GSTUMT00002663001	PAPS	BM271	**CccGGTACCggt**ATGACAGTAACTCCAACTGAATACC	58	927
GSTUMT00002663001	PAPS	BM272	**cctTCTAGA**CTAGATGGCCCATTGTAACC	57	927

Tabella 3: primer utilizzati per l'amplificazione delle sequenze clonate all'interno del vettore pB1H2ω5

Procedura per la costruzione della *library*

Per quanto riguarda la costruzione della libreria è necessario in primo luogo decidere la lunghezza della regione casuale che si desidera clonare nel vettore, generalmente compresa tra 18 e 28 nucleotidi. Maggiore è la lunghezza della sequenza maggiore sarà la probabilità di rappresentare i motivi di legame di un fattore di trascrizione che generalmente sono attorno alle 8-10 paia di basi ma sarà anche maggiore la difficoltà di interpretazione del motivo di legame. Per questo è consigliata una regione di 20 nucleotidi che rappresenta un buon compromesso tra la rappresentatività dei motivi e l'interpretazione dei risultati delle analisi bioinformatiche per la predizione del motivo.
Il vettore utilizzato per la costruzione della libreria è il pH3U3-MCS che contiene un sito multiplo di clonaggio a monte dei geni *HIS3/URA3* all'interno del quale va inserita la sequenza casuale. I geni *reporter* HIS3 e URA3 sono posti sotto il controllo di un promotore debole *lac* a cui si lega la RNA polimerasi batterica quando viene reclutata dal fattore di trascrizione. Il vettore possiede una origine di replicazione fagica f1 per cui all'interno della cellula è mantenuto a singola copia.
Per la costruzione della libreria contenente un inserto di 20nt a sequenza casuale si è seguito il protocollo descritto da Noyes (Noyes et al., 2008b) per cui vengono utilizzati 3 primer, uno lungo contenente 20 basi n fiancheggiate da regioni conservate a cui si andranno ad appaiare gli altri due primer di minor lunghezza in modo da creare due estremità protrudenti compatibili con il clonaggio EcoRI-NotI nel vettore pH3U3.



Gli oligonucleotidi devono essere fosforilati. Per la formazione della sequenza casuale si prepara una soluzione contenente tutti i primer a concentrazione finale 0,2 µM, si scaldano a 100°C per 10' e poi si fanno raffreddare in modo graduale per favorirne l'appaiamento. Il frammento così ottenuto viene messo in reazione di ligazione con 1 µg di vettore pH3U3 digerito EcoRI e NotI e lasciato overnight a 16°C. Attraverso prove di ligazione è stato possibile determinare che la quantità ottimale di frammento era di 8µl. La ligazione, dopo precipitazione con etanolo, viene rispospesa in 3µl e la metà viene utilizzata per trasformare un ceppo batterico One Shot® TOP10 Electrocomp™ E. coli. Le cellule batteriche sintetizzano il frammento di 20nt inserendo nucleotidi in modo casuale creando in questo modo diverse combinazioni nucleotidiche. Questo

metodo di clonaggio evita il passaggio di *primer extension* per la formazione delle sequenze casuali.

Amplificazione della library

- Dopo trasformazione la libreria è stata messa in 50ml di terreno YT2X contenente kanamicina e lasciata in rotazione per 1 ora a 37°C, una piccola aliquota è stata utilizzata per il calcolo del titolo iniziale
- I 50ml di coltura vengono aggiunti a 150ml di YT2X addizionato di antibiotico e lasciati in agitazione per 3 ore a 37°C in modo da ottenere una buona espansione della library. Dopo l'espansione una aliquota è stata utilizzata per il calcolo del titolo finale. La complessità della libreria deve essere compresa tra i 10^6-10^7 cloni per µg di DNA ligato e l'espansione deve essere di almeno 10 volte.
- Dopo le tre ore di espansione i 200ml di coltura vengono aggiunti a 1litro di YT2X con kanamicina e lasciati in agitazione a 37°C overnight. Questo passaggio è indispensabile per poter ottenere una sufficiente quantità di DNA dato che il vettore pH3U3 è a singola copia all'interno della cellula e si è notato che l'estrazione plasmidica con l'utilizzo di un kit commerciale non è abbastanza efficiente.
- Si procede con l'estrazione della library mediante un procedimento manuale di lisi alcalina su metà del pellet ottenuto dalla coltura overnignt. La restante metà viene conservata a -80°C.

Protocollo di lisi alcalina manuale

- Divisione della coltura in biberon e centrifugazione per 15' a 5000rpm
- Lavaggio delle cellule con 100ml di TEN e centrifugazione per 15' a 5000rpm
- Il pellet contenuto in ogni biberon viene trattato con 4ml di soluzione I, poi si aggiungono 8ml di soluzione II (0,1% SDS+ 0,2N NaOH) , si mescola per inversione e si incubano per 3' a temperatura ambiente e successivamente si aggiungono 6ml di soluzione III e si mescola vigorosamente ma senza vortex.
- Centrifugazione per 30' a 5000rpm a 4°C
- Si preleva il surnatante, si aggiungono 0,6 volumi di isopropanolo e 40µl di glicogeno e si centrifugano a 15000rcf a 20°C per 30'
- Si lascia asciugare bene il pellet e si risospende in 1ml di H_2O

- Precipitazione con etanolo e sodio acetato
- Purificazione su colonna utilizzando il kit Promega e risospensione in 30µl di H_2O.
- Quantificazione su gel di agarosio.

Protocollo per la contro-selezione

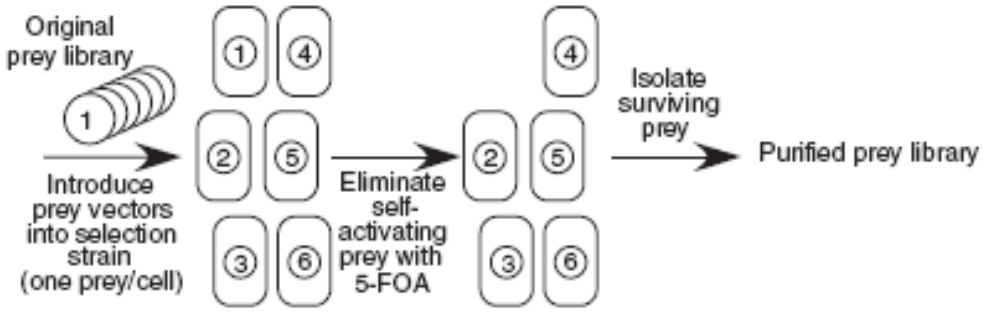


Figura 4: procedura di selezione negativa (Meng et al., 2005)

- Trasformare due aliquote del ceppo USO (4OD per trasformazione) con 100ng della libreria. Risospenderle in 1ml di SOC, unirle in un unico falcon e mettere in agitazione a 37°C per 1h.
- Centrifugare le cellule e risospenderle in 10ml di terreno minimo YM addizionato di kanamicina e metterle a 37°C in agitazione per 1h.
- Lavare le cellule eseguendo due lavaggi in H_2O e risospenderle in 1ml di terreno YM. Una aliquota viene utilizzata per la titolazione mentre la restante coltura viene conservata a 4°C.
- Noto il titolo piastrare con l'ausilio di biglie di vetro un totale di $2*10^7$ colonie suddivise in 40 piastroni da 50ml di terreno selettivo YM contenente 5-FOA 2,5mM e si lasciare crescere a 37°C per 12-13 h al massimo per cercare di ridurre il più possibile la crescita dei falsi positivi.
- Raccolta delle colonie cresciute utilizzando 10ml di H_2O per ogni piastra e calcolo del titolo
- Ripetere la piastratura su 5-FOA per migliorare l'eliminazione degli auto-attivatori.
- Raccolta delle colonie e preparazione plasmidica mediante lisi alcalina con protocollo manuale.

Si è reso necessario un secondo passaggio di contro-selezione perché si è notato che la prima selezione su 5-FOA non risultava sufficiente per eliminare gli auto-

attivatori poiché, nonostante molteplici prove, si è riscontrata una ridotta crescita delle colonie su 5-FOA ma non la morte di parte delle colonie piastrate come atteso.

Selezione positiva

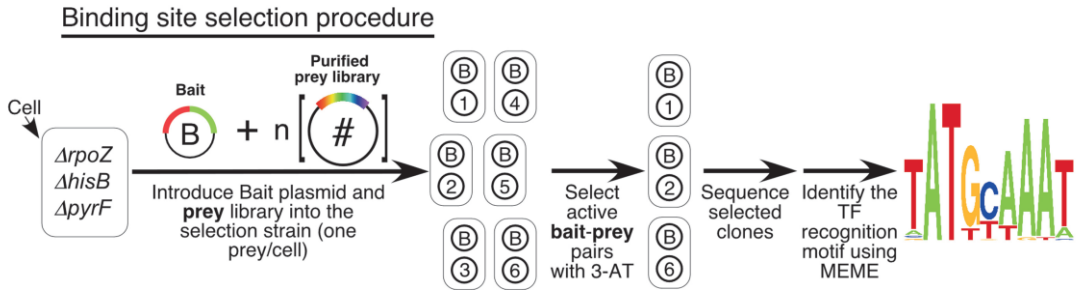


Figura 5: procedura di selezione positiva (Noyes et al., 2008b)

Una volta ottenuta la libreria ripulita si procede con la selezione dei cloni contenenti le sequenze riconosciute da un determinato fattore di trascrizione di interesse. Il ceppo USO in precedenza trasformato con il vettore pB1H2ω-TF, con il controllo positivo pB1H2ω2-Zif268 e con il controllo negativo pB1H2ω2-mutOdd viene trasformato con 100ng di *library* ripulita. Il controllo positivo è rappresentato da un DBD di un fattore di trascrizione per cui è nota la sequenza a cui si lega mentre il controllo negativo è il DBD di una proteina di *Drosophila* mutato in modo da perdere la capacità di legarsi al DNA. Nella selezione positiva, perché possa essere considerata attendibile, il rapporto tra colonie cresciute con il controllo positivo e negativo deve essere almeno di 10. Nel nostro caso il risultato migliore ottenuto è stato di 6 volte. Una volta eseguita la trasformazione le cellule vengono risospese in1ml di SOC e lasciate in agitazione a 37°C per 1h, poi centrifugate e risospese in terreno minimo NM contenente istidina allo 0,1%, kanamicina 25µg/ml, carbanecillina 100 µg/ml e IPTG 10µM e lasciati in agitazione a 37°C per 1h. Dopo si eseguono 2 lavaggio con H_2O e 2 lavaggi con terreno NM, si risospendono le cellule in terreno NM e se ne usa una aliquota per il titolo mentre il restante viene diviso in due piastre da 50ml di terreno NM contenente IPTG 10µM e 3AT 5 e 10mM.

Le piastre sono incubate a 37°C per 24-36 h. Le colonie cresciute rappresentano i cloni positivi, si procede allora con l'amplificazione del frammento della *library* e il suo sequenziamento. In realtà nel nostro caso abbiamo verificato che solo il 50% delle sequenze ottenute dal controllo positivo contenevano il motivo di legame del fattore di trascrizione Zif268. Questo dato non è sufficiente per poter ottenere una predizione del

motivo attendibile e significativa, per questa ragione sono stati aggiunti ulteriori passaggi al protocollo tradizionale.

Come suggeriscono gli autori del sistema in caso di fallimento della selezione si procede con la raccolta delle colonie cresciute su 3AT dalle quali si isolano i plasmidi. La miscela dei due plasmidi viene sottoposta a digestione con l'enzima di restrizione XmnI che digerisce solamente il vettore pB1H2 e con la reazione di digestione purificata si trasforma il ceppo USO mediante trasformazione chimica e si procede ad una sorta di ripulitura in piccolo. I trasformanti si risospendono in 1ml di SOC e si mettono in agitazione a 37°C per 1h. Poi le cellule vanno centrifugate e risospese in 1ml di terreno minimo YM addizionato di kanamicina e messe a 37°C in agitazione per 1h. Dopo si eseguono due lavaggi in H_2O e si risospendono in 1ml di terreno YM e si piastrano in una unica piastra da 25 ml di terreno selettivo YM contenente 5-FOA 2,5mM. In questo modo si dovrebbero eliminare i falsi positivi.

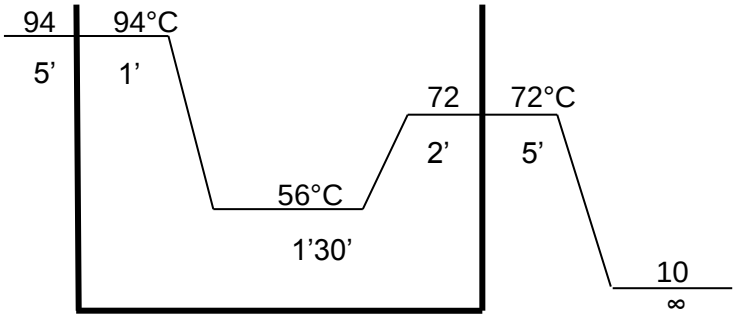
Visto i problemi avuti con la selezione su 5-FOA si è deciso di aggiungere un ulteriore passaggio di selezione, non ipotizzata nei diversi articoli scritti su questo metodo, che prevede la raccolta delle colonie crescite su 5-FOA in formato a 96 su piastre di terreno YT2X con kanamicina, una volta cresciute tutte le colonie sono replicate mediante l'utilizzo di un multipinner su terreno NM con 3AT 5 o 10mM in base alla concentrazione della piastra di selezione iniziale. Dato che il ceppo USO è stato trasformato con la miscela di digestione XmnI e che quindi possiede solo il vettore della *library* non dovrebbe essere in grado di crescere su 3AT perché in assenza del fattore di trascrizione non dovrebbe esserci attivazione dei geni *reporter* ad eccezione dei casi di auto-attivatori. Di conseguenza si considerano cloni positivi alla selezione del TF i cloni che non sono in grado di crescere in presenza di 3AT.

Le sequenze dei cloni positivi vengono isolate mediante PCR su colonia e inviate a sequenziare. Per ogni fattore di trascrizione si sequenziano dai 20 ai 40 prodotti di PCR in modo da poter identificare un motivo di legame al DNA significativo. Le sequenze vengono analizzate mediante il programma bioinformatico MEME che consente l'identificazione di un motivo conservato. Una volta identificato il motivo di legame del fattore di trascrizione è necessaria una ricerca di questo motivo all'interno dei promotori dei geni dell'organismo in modo da determinare quali sono i potenziali geni regolati dal TF. In questo modo si riesce a determinare la funzione del fattore di trascrizione e il suo meccanismo di azione.

Protocollo di colony PCR per l'amplificazione su vettore pH3U3

Dato che il vettore è a singola copia all'interno delle cellule batteriche si utilizzano condizioni di PCR diverse rispetto al protocollo standard.

Buffer 5X	10 µl
dNTPs 10mM	1,5 µl
primer Fw 10µM (BM200)	5 µl
primer Re 10µM (BM205)	5 µl
Go-Taq (5U/µl)	0.4 µl
H_2O	28,1 µl
	50 µl



35 cicli

I prodotti di PCR così amplificati vengono purificati con il metodo Millipore (già descritto in precedenza) e mandati a sequenziare presso la Macrogene Europe.

Analisi delle sequenze per l'identificazione del motivo conservato

Una volta ottenuti i risultati dei sequenziamenti vengono eliminate le sequenze che possiedono bassa qualità e vengono isolate le sequenze di 20 nt che corrispondono alla *library*. Queste cassette vengono poi analizzate attraverso il programma bioinformatico MEME (http://meme.sdsc.edu/meme/cgi-bin/meme.cgi) che permette di isolare un motivo conservato all'interno delle sequenze in esame. Questo programma permette di impostare alcuni parametri come la lunghezza del motivo che vogliamo isolare, quante volte potrebbe essere ripetuto all'interno delle sequenze analizzate, e quanti diversi motivi deve cercare all'interno delle sequenze.
Una volta identificato il motivo è possibile esportare da MEME la matrice di sostituzione posizione specifica (PSSM) che verrà utilizzata per analizzare le regioni promotrici a monte di tutti i geni codificati dal genoma per identificare le posizioni dei motivi di legame al DNA all'interno delle regioni regolatrici.

Risultati e discussione

1. Analisi *in silico* dei fattori di trascrizione codificati dal genoma di *Tuber melanosporum*

Il ciclo vitale degli ascomiceti simbiotici e in particolare di *Tuber melanosporum* è composto da tre fasi: micelio (FLM), ectomicorriza (ECM) e corpo fruttifero (FB). Poco sino ad ora si conosce per quanto riguarda i meccanismi che stanno alla base della transizione tra le diverse fasi del ciclo vitale. Alcune importanti informazioni sono state ottenute analizzando i profili trascrittomici che permettono l'associazione tra geni differenzialmente espressi e specificità di tessuto. Come evidenziato anche in basidiomiceti la fase micorrizica è spesso correlata con l'aumento dell'espressione di specifici gruppi di geni. Ad esempio risultano essere sovraespressi geni coinvolti nella divisione e proliferazione cellulare, nella sintesi della parete e della matrice extracellulare, nella risposta a stress e nel metabolismo primario (Martin et al., 2007).
L'analisi trascrittomica condotta per i funghi *T. melanosporum* e *L. bicolor* ha evidenziato una variazione dei profili di espressione di alcuni geni durante l'instaurarsi della simbiosi. Ora è necessario uno studio approfondito per determinare quali sono i fattori di trascrizione responsabili delle variazioni dell'espressione genica nelle diverse condizioni e soprattutto quali fattori regolano la transizione tra i diversi stadi del ciclo vitale, con particolare interesse per la fase simbiotica e per lo sviluppo del corpo fruttifero.
Il sequenziamento del genoma di *T. melanosporum* ha consentito di poter catalogare l'intero repertorio dei fattori di trascrizione codificati dal genoma. La classificazione *in silico* dei fattori di trascrizione è possibile mediante il riconoscimento, attraverso programmi bioinformatici, del dominio di legame al DNA (DBD) presente nella sequenza. Inoltre sono stati ricercati i putativi fattori di trascrizione presenti nel genoma di *T. melanosporum* omologhi a TF già caratterizzati in *S. cerevisiae*, *N. crassa* e *A. nidulans.*
I TF isolati sono stati catalogati nelle diverse famiglie in base alla tipologia del loro DBD. Come atteso ritroviamo le classi maggiormente rappresentate nei genomi fungini: bZIP, bHLH, e gli *zinc finger*, in particolare in *T. melanosporum* sono più abbondanti

fattori di trascrizioni appartenenti agli *zinc cluster*, unici dei funghi, e agli *zinc finger* C2H2.

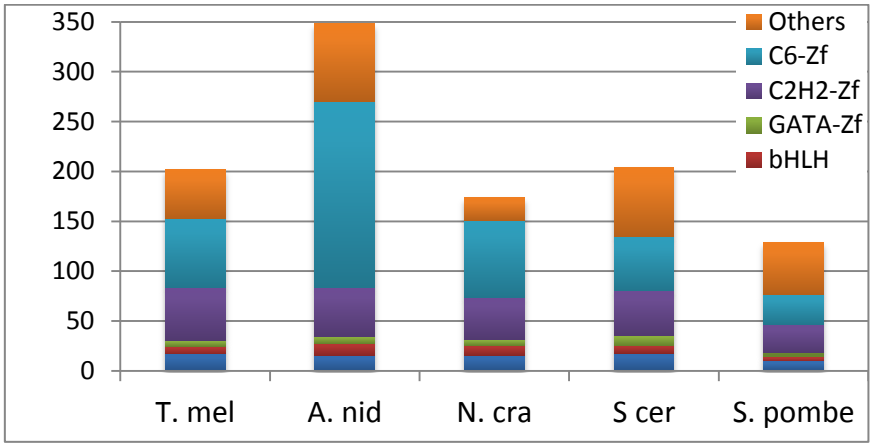


Fig1: Suddivisione dei fattori di trascrizione di diversi organismi nelle diverse classi. T.mel=*Tuber melanosporum;* A.nid= *Aspergillus nidulans;* N.cra=*Neurospora crassa;* S.cer=*Saccharomyces cerevisiae;* S.pombe=*Schizosaccharomyces pombe*

Il numero di fattori di trascrizione codificati dal genoma di *Tuber* è in linea con quello degli altri ascomiceti e anche la loro distribuzione nelle diverse classi non mostra differenze significative. L'unica eccezione è rappresentata dal confronto con *S. pombe* e *A. nidulans.* Quest'ultimo possiede un numero di TF doppio rispetto a *T.melanosporum* con una forte prevalenza di *zinc cluster*.

L'attribuzione di una funzione per omologia con proteine a funzione nota potrebbe portare a risultati ambigui, soprattutto nei casi in cui l'omologia di sequenza tra due fattori di trascrizione è ristretta al solo DBD e non al resto della proteina. Per questo si è reso necessario eseguire altre analisi BLAST eliminando dall'analisi il DBD in modo da evidenziare l'eventuale presenza all'interno della sequenza di altre regioni omologhe. In base a questi risultati i fattori di trascrizione sono stati suddivisi in tre gruppi: pTF se l'omologia a fattori di trascrizione con funzione caratterizzata era estesa lungo l'intera sequenza proteica, CH-TF se l'omologia era estesa a tutta la sequenza ma la funzione è sconosciuta e HP-DBD per i fattori unici di *Tuber* la cui omologia con altri TF è ristretta solo al dominio di legame al DNA.

Dall'analisi *in silico* si sono riscontrati 201 fattori di trascrizione suddivisi in 42 HP-DBD, 102 pTF e 57 CH-TF. Tutti i fattori di trascrizione identificati risultano essere espressi in almeno uno dei tre stadi del ciclo vitale. A 91 dei 102 pTF può essere assegnata una specifica funzione di controllo della trascrizione poiché simili a TF noti in altri funghi, in particolare in *N. crassa* e *A.nidulans.*

I pTF di *Tuber* possono essere suddivisi in cinque gruppi in base alla funzione dei loro ipotetici geni bersaglio: modificazione della parete, sviluppo, ciclo cellulare, metabolismo, risposta a stress o stimoli. (Tabella 1 in appendice)

In *Tuber* si ritrovano tutti i fattori di trascrizione, noti in altri organismi, che possiedono un ruolo nella regolazione di geni codificanti per enzimi coinvolti nella degradazione della parete cellulare, come ad esempio il repressore CreA, il repressore Ace1, il regolatore del pH PacC, il complesso Hap, il regolatore AreA, e l'attivatore dei geni che codificano per enzimi di degradazione della cellulosa/emicellulaosa XlnR. L'unica eccezione è il fattore Ace2 che sembra essere assente in *Tuber* ma è necessario sottolineare che questa proteina sembra mancare anche in *A. nidulans*, *A. niger* e *Magnaporte grisea* suggerendo che in questi funghi la funzione di Ace2 possa essere svolta da una altro fattore di trascrizione.

Questi enzimi degradativi sono indispensabili per il meccanismo di instaurazione della simbiosi, infatti sono responsabili della degradazione dei polimeri che compongono la parete cellulare vegetale come cellulosa, emicellulosa, lignina e pectina. L'espressione di questi enzimi è regolata da attivatori trascrizionali che ne consentono l'espressione in presenza di questi biopolimeri.

CreA è coinvolto nella repressione da catabolita, infatti in presenza di glucosio entra nel nucleo dove riconosce i promotori di cellulasi, emicellulasi e pectinasi e ne blocca l'espressione (Aro et al., 2005).

Il fattore di trascrizione XlnR, uno *zinc cluster*, regola l'espressione di endoxilanasi, xilosidasi, endoglucanasi, cellobioidrolasi e di alcune emicellulasi. Questo TF non è solo coinvolto nella regolazione di enzimi extracellulari ma anche nella regolazione di enzimi intracellulari coinvolti nel catabolismo dello xilosio. Questo dimostra la presenza nei funghi filamentosi di una connessione tra la regolazione di enzimi catabolici intra e extracellulari (Aro et al., 2005).

PacC regola l'espressione in base a cambiamenti di pH andando ad attivare l'espressione di geni quando la cellula si ritrova in un ambiente alcalino, tra i geni regolati da questa proteina si trovano anche enzimi che degradano la pectina (Aro et al., 2005).

AreA è un fattore di trascrizione coinvolto sia nella regolazione del metabolismo dell'azoto, ad esempio attiva i geni responsabili per l'utilizzo della prolina, sia nella regolazione di cellulasi suggerendo un legame tra l'utilizzo di fonti di carbonio e di azoto.

Il genoma di *Tuber* codifica anche per il fattore di trascrizione Ctf1 che regola l'espressione di una cutinasi, enzima secreto dalla cellula per favorire la penetrazione del fungo attraverso lo strato cuticolare della pianta (Rabanales, 2008).

Diversi omologhi di fattori di trascrizione di *N. crassa* coinvolti nello sviluppo si ritrovano in *Tuber* come per esempio quelli coinvolti nella crescita ifale e nello sviluppo sessuale/asessuale ma anche regolatori della formazione del corpo fruttifero e della risposta ai feromoni come ad esempio SteA ortologo del regolatore di lievito Ste12, StuA, NosA, Ste11 e la proteina del *mating type* Mat1-2-1.

Inaspettatamente Mat1-2-1 e SteA sono poco espressi nel corpo fruttifero mentre un ortologo di NsdD di *Aspergillus* che è noto essere coinvolto nello sviluppo sessuale è assente in *Tuber*. Sono presenti invece gli omologhi di altri fattori di trascrizione coinvolti nella formazione del corpo fruttifero o nella formazione dei conidi. Si ritrovano omologhi ai TF di *N. crassa* come le proteine White collar WC1 e WC2 che sono responsabili dell'attivazione di geni regolati dalla percezione della luce (Ballario and Macino, 1997).

È presente il regolatore CpcA, responsabile del controllo dell'espressione di geni in condizioni di carenza di amminoacidi nel terreno, che nei funghi filamentosi rappresenta l'omologo della proteina di lievito Gcn4 (Hoffmann et al., 2001).

Per quanto riguarda i regolatori dello sviluppo ritroviamo anche i fattori di trascrizione FlbB, FlbC e FlbD coinvolti nel meccanismo di regolazione che, a seguito della percezione di determinati stimoli ambientali, provoca l'attivazione di un *pathway* che determina l'arresto dell'estensione delle ife e la successiva formazione delle spore asessuali nei conidiofori (Etxebeste et al., 2008).

Il genoma codifica anche per i maggiori regolatori del metabolismo dell'azoto e del carbonio. Sono presenti, infatti, regolatori per l'utilizzo di forme di azoto alternative e attivatori di *pathway* specifici per il metabolismo di particolari fonti di azoto.

Dall'analisi si ritrovano anche regolatori della biosintesi di amminoacidi aromatici e dell'assimilazione dello zolfo che possono essere importanti per la formazione dei composti volatili che caratterizzano l'aroma del tartufo.

La maggior parte dei TF sono codificati da una singola copia del gene, le uniche eccezioni sono, tralasciando pseudogeni, rappresentate dall'omologo del regolatore VAD3 di *N. crassa* che è presente i due copie e l'attivatore del metabolismo del nitrato NirA che è presente in tre copie. NirA è responsabile della regolazione della nitrato e nitrito reduttasi e della nitrato permeasi che sono coinvolti nell'assimilazione del nitrato (Strauss et al., 1998). La presenza di tre geni codificanti per NirA nel genoma di *Tuber*

potrebbe essere potenzialmente correlata alla fase di simbiosi tra fungo e pianta poiché il metabolismo dell'azoto viene alterato durante l'instaurazione della simbiosi.

Fattori di trascrizione preferenzialmente espressi in ECM o FB

Dall'analisi dei profili trascrittomici dei fattori di trascrizione codificati dal genoma di *Tuber* si evidenziano 25 TF risultanti essere sovraespressi di almeno 5 volte nell'ectomicorriza (16) o nel corpo fruttifero (9) rispetto al micelio con valori che possono raggiungere una espressione di 49 volte per l'ECM e di 4500 volte per il FB. Tra questi sono presenti un omologo del regolatore della risposta a stress Rpn4, due CH-TF e uno *zinc cluster* specifico di *Tuber* risultano essere over-espressi più di 5 volte rispetto al micelio in entrambi gli stadi vitali.

Gene model	Name	DBD	ECM	FB	Putative function
GSTUMT00004347001	TmelXLNR	C6 Zn cluster			***Cell wall modification***
GSTUMT00012451001	**HP-DBD, TelXLNR2**	C6 Zn cluster			
GSTUMT00005078001	TmelABAA	TEA			***Development & hyphal growth***
GSTUMT00007803001	TmelDEVR	bHLH			
GSTUMT00008124001	TmelFF7	C6 Zn cluster			
GSTUMT00009849001	TmelFLBD	Myb-like factors			
GSTUMT00005459001	TmelHMS1	bHLH			
GSTUMT00008596001	**TmelACUM**	C6 Zn cluster			***Carbon metabolism***
GSTUMT00012254001	**TmelFARB**	C6 Zn cluster			
GSTUMT00000111001	TmelACUK	C6 Zn cluster			
GSTUMT00000814001	TmelCBF1	bHLH			***Sulfur metabolism***
GSTUMT00012172001	TmelMET32	C2H2			
GSTUMT00006696001	**TmelRPN4**	C2H2			***Stress response***
GSTUMT00004466001	TmelSFL1	HSF			
GSTUMT00008521001	TmelSFP1	C2H2			
GSTUMT00007914001	**CH-TF**	C2H2			***Conserved hypothetical transcription factors***
GSTUMT00006533001	CH-TF	C6 Zn cluster			
GSTUMT00009188001	CH-TF	C6 Zn cluster			
GSTUMT00011438001	CH-TF	C6 Zn cluster			
GSTUMT00004451001	**CH-TF**	Homeobox			
GSTUMT00006234001	**CH-TF**	CP2			
GSTUMT00000561001	CH-TF	C6 Zn cluster			
GSTUMT00004341001	CH-TF	C6 Zn cluster			
GSTUMT00009814001	**HP-DBD**	C6 Zn cluster			***Tuber specific DBD-containing hypothetical proteins***
GSTUMT00004776001	**HP-DBD**	Forkhead			
GSTUMT00005836001	**HP-DBD**	C6 Zn cluster			
GSTUMT00006134001	**HP-DBD**	C6 Zn cluster			
GSTUMT00011475001	HP-DBD	Myb-like factors			
GSTUMT00000279001	HP-DBD	C2H2			

Valori di espressione rapportati al micelio

< 5
5 :: 10
10 :: 100
> 100

Tabella1: Tabella dei fattori di trascrizione maggiormente espressi in ectomicorriza o in corpo fruttifero. Codice colore: giallo rapporto di espressione con il micelio compreso tra 5 e 10, arancio rapporto tra 10 e 100, rosso rapporto maggiore di 100.

Tra i fattori di trascrizione maggiormente espressi in ectomicorriza sono presenti i TF coinvolti nella modificazione della parete batterica XlnR e uno *zinc cluster* unico di *Tuber* simile a XlnR1 che abbiamo denominato TmelXlnR2 anche se non è ancora noto il suo meccanismo. Quest'ultimo risulta essere uno dei trascritti maggiormente espressi in ECM, inoltre XlnR1 risulta essere sovraespresso 5 volte rispetto al micelio mentre è 10 volte meno espresso rispetto a XlnR2.
Sono sovraespressi in ECM anche il regolatore della morfologia TmelHms1, il regolatore dello sviluppo asessuale TmelFlbD e TmelDevR e il regolatore della sviluppo sessuale TmelFf7. Risultano sovraespressi anche TmelSFP1 e TmelSFL1 suggerendo che la regolazione in risposta a stress e la differenziazione morfologica siano processi coinvolti nella formazione dell'ectomicorriza. Sono presenti anche fattori a funzione sconosciuta sia appartenenti ai CH-TF che agli HP-DBD.
Nove fattori di trascrizione sono sovraespressi nel corpo fruttifero, di cui 6 appartengono a *zinc cluster* mentre 5 possiedono un omologo a funzione nota. Ritroviamo TmelAbaA un regolatore dello sviluppo asessuale, TmelAcuM e TmelAcuK attivatori della gluconeogenesi, TmelCbf1 regolatore del metabolismo dello zolfo e TmelFarB un attivatore del metabolismo degli acidi grassi a catena corta. L'espressione preferenziale di questi geni è in accordo con il fatto che la gluconeogenesi e l'assimilazione dello zolfo sono *pathway* metabolici altamente supportati nel corpo fruttifero. Tra le altre 4 proteine maggiormente espresse ritroviamo due CH-TF e due HP-DBD, di cui uno risulta essere il secondo trascritto maggiormente espresso nel FB. Nessun TF coinvolto nella regolazione dello sviluppo sessuale risulta sovraespresso nel corpo fruttifero.

2. *Screening* genomico funzionale degli attivatori trascrizionali di *Tuber melanosporum*

Per validare funzionalmente alcuni dei fattori di trascrizione predetti e per scoprire nuovi attivatori della trascrizione alle analisi bioinformatiche è seguito il *transcriptional activator trap*, uno *screening* condotto in lievito che permette l'isolamento degli attivatori della trascrizione.
Per identificare gli attivatori trascrizionali espressi nelle tre fasi del ciclo vitale di *T.melanosporum*, sono state trasformate le genoteche F2BD di corpo fruttifero (FB), M1BD e M2BD di micelio (FLM) e E1BD di ectomicorriza (ECM) in lievito *S. cerevisiae*.
La genoteca è stata ingegnerizzata tramite fusione di cDNA di *Tuber melanosporum* con il dominio di legame al DNA dell'attivatore trascrizionale di lievito GAL4 ed inserite

nel vettore di espressione pDEST32. Le genoteche F2BD, M2BD e E1BD possiedono inserti di taglia media di 1kb mentre la M1BD di 1,3 kb.

Il ceppo di lievito Mav103 utilizzato per la trasformazione contiene i geni *reporter HIS3, LacZ* e *URA3*. Questi tre geni posseggono a monte una regione regolativa che contiene la UAS di GAL4 che quindi è riconosciuta dal *DNA binding domain* di GAL4. La trascrizione dei geni *reporter* è attivata solo se il cDNA di *Tuber* possiede la funzione di attivatore trascrizionale.

Per la trasformazione del lievito sono stati utilizzati 3µg della genoteca F2BD, 10µg ciascuna delle M1BD e M2BD e 10µg della E1BD e i trasformanti di ogni genoteca sono stati piastrati su 40 piastre contenenti 50 ml di terreno minimo privo di istidina e leucina, addizionato di 3-AT alla concentrazione 25mM in modo da poter selezionare trasformanti in grado di attivare la trascrizione del gene *reporter HIS3*.

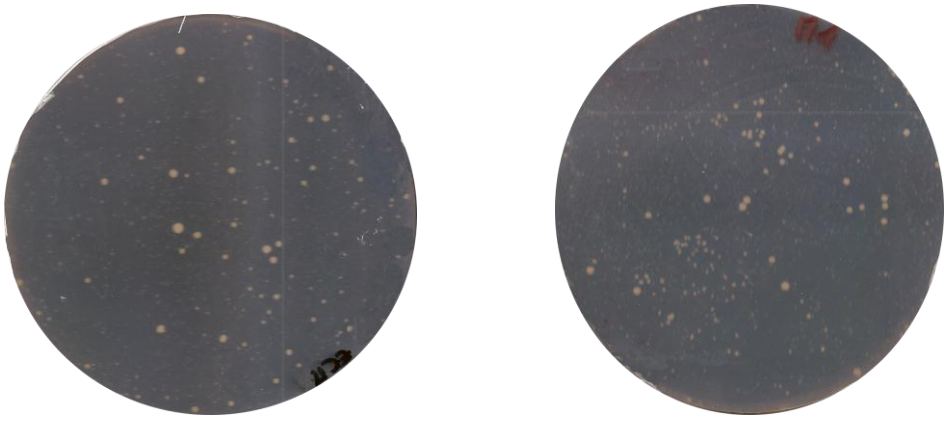


Fig2: Piastre SD –Leu –His 3AT 25mM della trasformazione del ceppo di lievito Mav103. La prima con la genoteca E1BD mentre la seconda con quella M1BD.

Il numero di colonie piastrate totali è stato di $5*10^5$ per le genoteche F2BD e E1BD mentre di $3*10^5$ e $2,2*10^5$ per M1BD e M2BD rispettivamente. Le colonie cresciute maggiormente sono state prelevate dalle piastre di trasformazione e fatte crescere in piastre SD-Leu in formato a 96 pozzetti per un giorno a 30°C. Sono state isolate 768 colonie per F2BD, 719 per M1BD, 370 per M2BD e 812 per E1BD. In ogni piastra sono stati aggiunti anche i controlli wt, m1, m2, pDEST32. Sono state assemblate piastre da 384 o da 96 colonie per ogni genoteca e queste piastre sono state il punto di partenza per effettuare i tre *screening* per valutare l'espressione dei tre geni *reporter*.

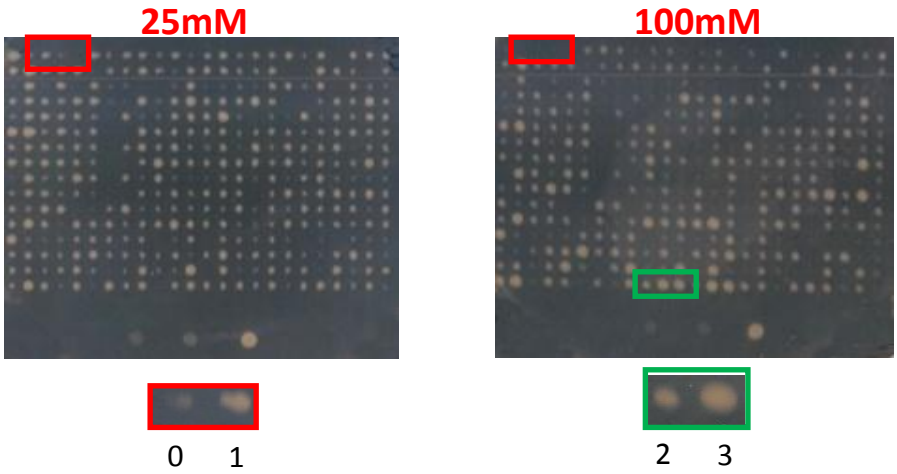


Fig3: Gene *reporter HIS3 v*alutazione del fenotipo in base alla dimensione di crescita delle colonie su terreno selettivo SD –Leu –His 3AT 25 e 100 mM

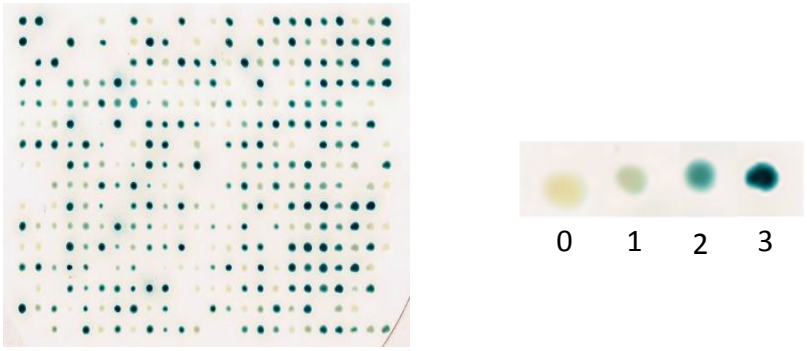


Fig4: Saggio *LacZ*, le colonie vengono classificate in base alla colorazione

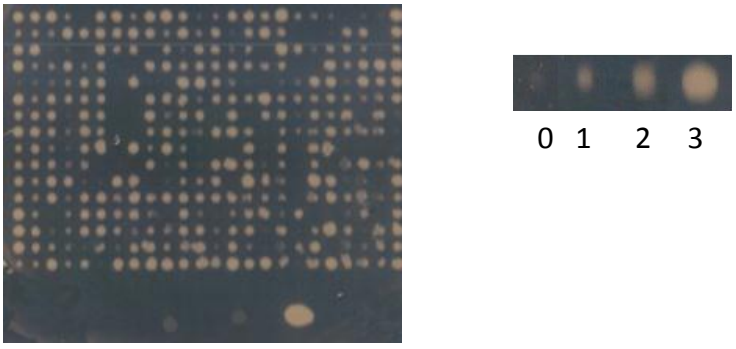


Fig5: Gene *reporter URA3* valutazione del fenotipo in base alle dimensioni di crescita delle colonie su terreno selettivo SD -Ura

Analisi delle sequenze positive allo *screening* TAT

Per la scelta dei cloni più appropriati da sequenziare, i valori assegnati ad ogni piastra sono stati analizzati tramite due criteri:

- sono state considerate le colonie con valore di espressione del gene *reporter* maggiore a 0. In questo modo si evidenziano tutte le colonie che attivavano la trascrizione almeno in modo lieve nei tre geni *reporter*;
- sono state selezionate tutte le colonie con almeno due geni *reporter* su tre con attivazione media.

Sono state quindi isolate le colonie che risultavano positive a entrambi i criteri: 205 colonie per F2BD (30% degli analizzati), 539 per M1BD (75%), 250 per M2BD (68%) e 414 per E1BD (50%) che sono state analizzate mediante PCR su colonia di lievito per ottenere gli ampliconi per il sequenziamento.

Sono risultate possedere buona qualità di sequenziamento un totale di 438 sequenze così suddivise:

- 152 dalla *library* di ectomicorriza;
- 160 dalla *library* di micelio;
- 126 dalla *library* di corpo fruttifero.

Dopo le analisi bioinformatiche queste sequenze sono state organizzate in 53 *contigs* e 88 *singletons,* per un totale di 141 sequenze uniche di cui 100 di *Tuber* e 41 di pianta. Le sequenze di pianta, come atteso, provengono dall'analisi della genoteca di ectomicorriza poiché è lo stadio di simbiosi.

Delle 100 sequenze uniche di *Tuber*, 37 sono rappresentate da fattori trascrizionali poichè contengono un *DNA binding domain* mentre le restanti 63 sequenze positive mancano di DBD noto. Tra queste si ritrovano 20 sequenze che chiamiamo "*Unconventional activators*" (UAs) simili a geni che esprimono proteine con localizzazione nucleare (8 sequenze) o mista nucleare/citoplasmatica (12 sequenze) per le quali un ruolo diretto o indiretto nella regolazione della trascrizione è stato precedentemente dimostrato.

Le restanti 43 sequenze le abbiamo denominate "*Putative unconventional activators*" (PUAs) poiché in letteratura non è noto un loro ruolo nella trascrizione.

	UNISEQU	CONTIG	SINGLETONS	TOTALE	ECM	FB	FM
Uas	20	9	11	41	4	22	15
PUAs	43	11	32	69	9	35	25
TFs	37	21	16	232	43	69	120
TOT. FUNGO	100	41	59	342	56	126	160
PIANTA TFs	22	7	15	68	68		
PIANTA PUas	6	4	2	14	14		
PIANTA UFs	13	1	12	14	14		
TOT. PIANTA	41	12	29	96	96		
TOT. PIANTA+FUNGO	141	53	88	438	152	126	160
37= 9 CHP, 6 HP, 22 TF							

Tabella2: Analisi delle sequenze di *Tuber*

Dall'analisi delle 37 sequenze che posseggono un dominio di legame al DNA ritroviamo 9 HP-DBD, 6 CH-TF e 22pTF. Questo ci ha permesso di validare funzionalmente il 25% degli HP-DBD trovati con l'analisi *in silico*, il 10,3% dei CH-TF e il 22% dei pTF. Considerando invece il totale dei fattori di trascrizione positivi al TAT abbiamo validato un quinto dei fattori di trascrizione codificati dal genoma di *Tuber.* Inoltre il TAT ha consentito l'identificazione di quattro TF che non erano stati isolati dalle analisi bioinformatiche poiché il *gene model* predetto dopo il sequenziamento mancava di un DBD riconoscibile poiché o era stato predetto in modo scorretto il primo esone o era presente solo una parte del DBD.

Tra i fattori funzionalmente validati che possiedono omologia con proteine già caratterizzate in altri funghi ritroviamo due regolatori della parete cellulare TmelAce1 e TmelMrgA, 4 TF che regolano lo sviluppo asessuale e la crescita ifale TmelTah1, TmelCol27, TmelCol21 e TmelKal1, 3 regolatori del ciclo cellulare TmelFkh2, TmelSep1 e TmelRme1, sette regolatori del metabolismo, TmelFacB che attiva geni per il metabolismo dell'acetato e TmelAcuM che regola la gluconeogenesi, TmelMeaB, TmelCpcA e TmelNirA1 coinvolti nel metabolismo dell'azoto, TmelMetR che regola il metabolismo dello zolfo e TmelFarB coinvolto nel metabolismo dei lipidi. Vengono validati funzionalmente anche sei attivatori che mediano la risposta a stress e stimoli ambientali TmelFcr3, TmelRpn4, TmelSp1, TmelHaa1, TmelHacA e TmelSltA.

Tra questi pTF ritroviamo 4 cDNA che vengono isolati in più di 20 cloni indipendenti: TmelCpcA (49 cloni), TmelMetR (25 cloni), TmelMeaB (27 cloni) e TmelKal1 (22 cloni)

inoltre rientra in questa categoria anche il fattore di trascrizione di pianta RAP2.12 per cui vengono isolati 24 cloni.

Sequence information			BlastP result			Functional analysis				
Gene model	**DBD**	**Gene name**	**Code**	**gene name**	**e-value**	**seq ID**	**ECM**	**FB**	**FLM**	**tot**
DBD-containing hypothetical proteins (HP-DBD)										
GSTUMT00003475001	Homeobox	GSTUMT00003475001	AN1217		2,00E-17	F2BD-D1_C12	0	1	0	1
GSTUMT00006134001	C6 *zinc cluster*	GSTUMT00006134001	AN8590		4,00E-05	contig56	1	3	2	6
GSTUMT00007102001*	C6 *zinc cluster*	GSTUMT00007102001	AN3433		5,00E-16	E1BD-G2_F04	1	0	0	1
GSTUMT00008508001	bZIP	GSTUMT00008508001	AN0951		8,00E-31	Contig22	5	0	4	9
GSTUMT00009205001	Myb-like factors	GSTUMT00009205001	NCU09197		3,00E-16	M2BD-N4_F08	0	0	1	1
GSTUMT00009240001*	bZIP	GSTUMT00009240001	nd			F2BD-D2_F03	0	1	0	1
Conserved hypothetical transcription factors (CH-TF)										
GSTUMT00002384001	C6 *zinc cluster*	GSTUMT00002384001	AN3650		3,00E-110	F2BD-S3_F01	0	1	0	1
GSTUMT00003655001	C6 *zinc cluster*	GSTUMT00003655001	NCU01478		4,00E-36	F2BD-S2_G09	0	1	0	1
GSTUMT00003728001	C2H2	GSTUMT00003728001	AN1437		1,00E-84	Contig39	0	4	1	5
GSTUMT00004065001	C6 *zinc cluster*	GSTUMT00004065001	AN2852		5,00E-44	Contig26	0	0	5	5
GSTUMT00005291001	Copper fist	GSTUMT00005291001	AN0658		6,00E-05	F2BD-D1_F05	0	1	0	1
GSTUMT00005306001*	C2H2	GSTUMT00005306001	NCU09807		8,00E-05	F2BD-S1_G10	0	1	0	1
GSTUMT00007914001	C2H2	GSTUMT00007914001	AN9176		1,00E-22	Contig47	0	2	0	2
GSTUMT00008944001	C6 *zinc cluster*	GSTUMT00008944001	AN7434		3,00E-11	Contig38	0	3	0	3
GSTUMT00010565001*	C2H2	GSTUMT00010565001	AN7583		2,00E-25	M1BD-N2_D05	0	0	1	1
Putative transcription factors (pTF)										
GSTUMT00000577001	bZIP	TmelCPC1	AN3675		4E-11	Contig20	2	0	47	49
GSTUMT00000600001	bZIP	TmelAP1	NP_013707	Yap1	1E-13	Contig35	0	2	7	9
GSTUMT00001766001	C6 *zinc cluster*	TmelCOL21	NCU03489	*col-21*	3E-37	Contig52	0	3	0	3
GSTUMT00002176001	C6 *zinc cluster*	TmelCOL27	NCU08651	*col-27*	4E-106	F2BD-D2_A05	0	1	0	1
GSTUMT00002283001	C6 *zinc cluster*	TmelNIRA-1	AN0098	nirA	2E-165	Contig24	0	0	11	11
GSTUMT00002563001	bZIP	TmelFCR3	Q8X229		1E-14	Contig23	0	0	3	3
GSTUMT00003085001	bZIP	TmelMEAB	AN4900		2E-36	Contig41	0	6	21	27
GSTUMT00003504001	C2H2	TmelRME1	NP_011558	Rme1	3E-06	E1BD-G1_B10	1	0	0	1
GSTUMT00004068001	Forkhead/Winged helix	TmelFKH2	NP_014331	Fkh2	9E-31	F2BD-D2_G04	0	1	0	1
GSTUMT00005465001	C6 *zinc cluster*	TmelFACB	AN0689	facB	3E-155	Contig48	0	2	0	2
GSTUMT00006159001	Forkhead/Winged helix	TmelSEP1	NP_596301.1		3E-34	M2BD-N3_C01	0	0	1	1
GSTUMT00006696001	C2H2	TmelRPN4	NP_010264	Rpn4	5E-12	Contig10	9	0	0	9
GSTUMT00006959001	Copper fist	TmelHAA1	NP_015333	Haa1	3E-13	Contig12	3	1	0	4
GSTUMT00007706001	Homeobox	TmelKAL1	NCU03593	*kal-1*	2E-65	Contig31	19	1	2	22
GSTUMT00008596001	C6 *zinc cluster*	TmelACUM	AN6293		1E-117	Contig51	0	3	0	3
GSTUMT00008609001	bZIP	TmelMETR	AN4361	metR	8E-27	Contig46	0	16	9	25
GSTUMT00009306001	C6 *zinc cluster*	TmelTAH1	NCU00289	*tah-1*	3E-75	E1BD-G2_B07	1	0	0	1
GSTUMT00009612001	C2H2	TmalSLTA	AN2919	sltA	3E-06	Contig32	0	0	2	2
GSTUMT00009625001	bZIP	TmelHACA	AN9397		7E-12	contig57	0	0	2	2
GSTUMT00009963001	C2H2	TmelMRGA	XP_002150457.1		6E-39	E1BD-G1_D09	1	0	0	1
GSTUMT00010551001	C2H2	TmelACE1	EAL92775.1		6E-31	M2BD-N3_F06	0	0	1	1
GSTUMT00012254001	C6 *zinc cluster*	TmelFARB	AN1425		3E-153	Contig43	0	15	0	15

Tabella 3: Fattori di trascrizione validati mediante TAT

Tra gli UAs abbiamo trovato varie chinasi, enzimi coinvolti nel processo di ubiquitinazione, proteine endocitotiche, actina ed *actin-related protein*.

L'actina è il maggior componente del citoscheletro ed è coinvolta in diversi processi cellulari tra cui mobilità, mitosi, trasporto intracellulare, endocitosi e secrezione, ma sono note anche sue azioni nel nucleo coinvolte nella trascrizione, rimodellamento della cromatina e esporto di RNA dal nucleo. Può avere un ruolo diretto nella trascrizione; infatti la β-actina è stata identificata come componente del complesso di pre-inizio dell'RNA polimerasi II. È stata descritta anche in letteratura per essere strettamente associata con la RNA polimerasi III tramite interazioni proteina-proteina. L'actina interagisce anche con differenti componenti del macchinario della RNA polimerasi I per la trascrizione dei geni ribosomiali. È stato scoperto che l'actina nucleare è una componente integrante dei complessi di rimodellamento della cromatina come ad esempio i complessi BAF nei mammiferi, BAP in *D. melanogaster complex* e Ino80 *complex* in *S. cerevisiae*. Per quanto riguarda le *actin related protein*, in particolare per l'α-actinina ritrovata nella nostra analisi funzionale, ci sono evidenze sperimentali di un suo ruolo come co-attivatore dei recettori androgenici nucleari (Zheng et al., 2009).

Alcune proteine endocitotiche posseggono un ruolo anche nel nucleo dove possono essere coinvolte nella regolazione della trascrizione genica. Come co-attivatori o co-repressori possono reclutare il macchinario trascrizionale (Pilecka et al., 2007). Due proteine endocitotiche descritte in letteratura con localizzazione nucleare sono state trovate anche tra i cloni TAT positivi: TmelEde1, che è omologa della proteina umana Eps15p (Vecchi et al., 2001) e TmelVps60, l'omologo della "*Charged multivesicular body protein* 5" (CHMP5). Eps15p è stata localizzata anche nel nucleo e visto che in risposta ad alcuni fattori di crescita viene fosforilata, è stato ipotizzato un suo ruolo come trasduttore del segnale, può entrare nel nucleo e regolare l'espressione genica (Poupon et al., 2002). Per CHMP5, da studi di doppio ibrido è emerso che interagisce con proteine coinvolte nel rimodellamento della cromatina e partecipa al processo di sumoilazione nel nucleo (Tsang et al., 2006).

La calmodulina è una proteina che lega il calcio ed è indispensabile all'interno della cellula poiché regola molti processi tra cui la risposta a diversi stress. Questa proteina trasduce il segnale andando a regolare processi sia citoplasmatici che nucleari come cambiamenti nella struttura del citoscheletro, attività metabolica ed espressione genica (Thorogate and Török, 2004). Inoltre la calmodulina media l'importo nel nucleo in modo

Ca2+ dipendente di molte proteine compresi fattori di trascrizione (Hanover et al., 2009).

Alcune chinasi sono note avere un ruolo nella trascrizione poiché regolano e interagiscono con fattori di trascrizione. Ad esempio la proteina Hog1 è indispensabile nel meccanismo di trasduzione del segnale in risposta a stress ed è in grado di modulare l'espressione genica fosforilando fattori di trascrizione. È stato inoltre dimostrato che si lega a determinati promotori e che il suo legame media il reclutamento, attraverso interazioni fisiche, di specifici fattori di trascrizione (Nadal and Posas, 2010).

La proteina Nas6 fa parte del complesso 19S del proteasoma, è stata localizzata nel nucleo e inoltre ci sono evidenze sperimentali del ruolo dell'intero complesso 19S nella regolazione dell'espressione genica. In particolare questo complesso sembra avere un ruolo nel formare il complesso di pre-inizio della trascrizione a livello di alcuni promotori di geni che vengono regolati in risposta a stress (Malik et al., 2009).

Sequence information						BLAST result			Sc	NLS
seq ID	ECM	FB	FLM	tot	Gene model	acc #	BLAST result	e-value		
Nuclear proteins and unconventional activators (UAs)										
Contig09	4	2	1	7	GSTUMT00006520001	EDV09130.1	serine/threonine-protein kinase KSP1	1E-47	Ksp1	Y
Contig29	0	0	2	2	GSTUMT00006162001	BAA23208.1	actin	6E-20	Act1	N
Contig30	0	0	4	4	GSTUMT00001950001	XP_002147958.1	low-temperature viability protein ltv1	9E-73	Ltv1	Y
Contig36	0	2	1	3	GSTUMT00009307001	EDN64569.1	EH domains and endocytosis-related protein	3E-49	Ede1	Y
Contig40	0	1	1	2	GSTUMT00005852001	NP_013214.1	Mitogen-activated protein kinase	7E-150	Hog1	N
Contig42	0	2	0	2	GSTUMT00009762001	EEH34348.1	DNA-directed RNA polymerase II subunit RPB1	0	Rpo21	N
Contig45	0	2	1	3	GSTUMT00011744001	XP_001934547.1	E3 ubiquitin-protein ligase NEDD4	0	Rsp5	Y
Contig54	0	4	0	4	GSTUMT00004791001	XP_001940596.1	peptide methionine sulfoxide reductase	8E-74	Mxr1	N
Contig55	0	3	0	3	GSTUMT00006491001	EEQ88820.1	alpha-actinin	0		N
F2BD-D1_C10	0	1	0	1	GSTUMT00004358001	XP_002173970.1	cryptic loci regulator 2	0		Y
F2BD-D2_C01	0	1	0	1	GSTUMT00012039001	XP_001931756.1	charged multivesicular body protein 5	2E-85	Vps60	N
F2BD-D2_E01	0	1	0	1	GSTUMT00005390001	XP_002378994.1	mitogen-activated protein kinase MAF1	8E-63	Maf1	Y
F2BD-S1_B12	0	1	0	1	GSTUMT00003661001	XP_002486373.1	phospholipase D1 (PLD1), putative	0	Spo14	Y
F2BD-S1_C03	0	1	0	1	GSTUMT00005490001	EDV12592.1	mitosis inhibitor protein kinase SWE1	2E-56	Swe1	Y
F2BD-S1_G11	0	1	0	1	GSTUMT00007890001	XP_002620593.1	protein vts1	4E-150	Vts1	Y
M1BD-N1_G04	0	0	1	1	GSTUMT00005863001	EEH06526.1	DASH complex subunit ask1	1E-56	Ask1	Y
M2BD-N3_A11	0	0	1	1	GSTUMT00010275001	XP_002484900.1	calmodulin	4E-49	Cmd1	N
M2BD-N3_G03	0	0	1	1	GSTUMT00003500001	XP_001329422.1	ankyrin repeat protein	7E-48	Nas6	N
M2BD-N4_B02	0	0	1	1	GSTUMT00000718001	XP_001263780.1	ubiquitin thiolesterase (OtuB1), putative	6E-61		N
M2BD-N4_D10	0	0	1	1	GSTUMT00011807001	EEH45709.1	casein kinase II beta 2 subunit	9E-37		N

Tabella 4: Proteine appartenenti alla categoria degli *unconventional activator* (Sc = omologhi in *S.cerevisiae*)

I "*Putative unconventional activators*" (PUAs) sono più problematici perché solo la metà di queste proteine possiede un segnale di localizzazione (NLS) predicibile *in silico*. Le restanti proteine potrebbero essere falsi positivi, dato che il sistema forza l'espressione della proteina nel nucleo, oppure avere un reale ruolo nel nucleo. In questo caso potrebbero rappresentare delle ipotetiche *moonlighting proteins* dato che, come evidenziato precedentemente, ci sono numerosi esempi di proteine non convenzionali per cui è stata dimostrata una funzione nella trascrizione.

Tra le PUAs ritroviamo 12 proteine di cui non si conosce la funzione di cui solo la metà presenta un segnale di localizzazione nucleare e 9 proteine uniche di *Tuber* tra cui solo due presentano un NLS. Tra quelle omologhe a proteine con una funzione predetta ritroviamo sei enzimi, due perossine coinvolte nella formazione del perossisoma, la calcineurina che è una fosfatasi, chinasi e proteine coinvolte nel trasporto vescicolare.

Data la peculiarità della classe PUAs sono necessari ulteriori studi per poter attribuire loro una nuova funzione o per catalogarle come falsi positivi. Il primo passo per verificare l'attendibilità di tali risultati è quello di eseguire esperimenti volti a verificare la reale localizzazione di queste proteine. Per questo motivo alcune di queste sono state sottoposte al *Nuclear Transportation Trap* (NTT).

seq ID	ECM	FB	FLM	tot	acc #	BLAST result	e-value	Sc
Contig33	0	1	1	2	EEY14547.1	Pex19 family protein	7E-48	Pex19
Contig34	0	1	6	7	XP_002144764.1	stromal membrane-associated protein	2E-101	Age2
Contig44	0	1	1	2	XP_001555369.1	calcineurin regulatory subunit B	2E-83	Cnb1
Contig49	0	3	0	3	XP_002625962.1	cAMP-dependent protein kinase *pathway* protein	9E-12	
contig58	0	1	1	2	XP_002562874.1	3-phosphoadenosine-5-phosphosulfate reductase	1E-111	Met16
E1BD-G1_H10	1	0	0	1	XP_960686.1	clock-controlled protein 6	0,0000005	
E1BD-G2_F02	1	0	0	1	EEQ27834.1	mucin	4E-32	
F2BD-D1_A09	0	1	0	1	NP_015080.1	Protein with a role in UDP-galactose transport	8E-54	Hut1p
F2BD-D1_B10	0	1	0	1	ZP_01885673.1	2-methylcitrate dehydratase	0	Pdh1
F2BD-D1_F07	0	1	0	1	AAT85969.1	Frp1	3E-34	
F2BD-D1_G08	0	1	0	1	XP_002624920.1	peroxin 20	5E-25	
F2BD-S1_A02	0	1	0	1	XP_001933661.1	ADP-ribosylation factor-binding protein GGA1	1E-162	Gga2
F2BD-S1_B08	0	1	0	1	EEH42280.1	autophagy-related protein 17	5E-61	Atg17
F2BD-S1_C06	0	1	0	1	XP_002172257.1	hamartin	1E-38	
M1BD-N2_H11	0	0	1	1	ACJ66700.1	sensor histidine kinase	5E-22	
M2BD-N3_B02	0	0	1	1	ACB47221.1	microtubule-associated protein	4E-59	Atg8
M2BD-N3_C11	0	0	1	1	XP_002487986.1	glucose-6-phosphate 1-dehydrogenase	0	Zwf1
M2BD-N3_G11	0	0	1	1	XP_756047.1	MAPKKK cascade protein kinase regulator Ste50	1E-83	Ste50
M2BD-N4_B10	0	0	1	1	XP_001933355.1	vacuolar protein-sorting protein bro1	0	Bro1
M2BD-N4_D08	0	0	1	1	XP_001275297.1	AMP-binding enzyme (Fum16), putative	0	Faa1
M2BD-N4_E03	0	0	1	1	NP_013006.1	NAD-dependent 5,10-methylenetetrahydrafolate	3E-99	Mtd1

Tabella 5: Proteine appartenenti ai *Putative Unconventional activator* che possiedono un omologo a funzione nota.

Delle 41 sequenze uniche di pianta, 22 possiedono DBD riconosciuto tra cui 12 hanno similarità di sequenza con fattori trascrizionali coinvolti nell'interazione simbionte. Le restanti sono formate da 18 PUAs ed una sequenza UA. (Tabella 2 in appendice)

I fattori di trascrizione contenenti un dominio NAC possono avere un ruolo nella risposta a stress biotici come l'infezione da patogeno, nella risposta a stress abiotici come shock termici e disidratazione (Olsen et al., 2005). Ci sono diversi esempi di fattori di trascrizione di pianta appartenenti a tale famiglia per cui è stato determinato un ruolo nella difesa da patogeni, uno è rappresentato dal TF chiamato TIP che funge da recettore virale, interagisce con una proteina del capside e attiva l'espressione di geni per la resistenza (Ren et al., 2000).

I fattori di trascrizione ERF (*ethylene-responsive-factor*) sono responsabili dell'espressione genica in risposta a freddo, siccità, patogeni o a trattamenti con etilene, acido salicilico o altre molecole (Singh, 2002).

I fattori di trascrizione con un dominio MYB rappresentano una delle maggiori famiglie di TF in pianta e sono coinvolti in numerosi processi come la morfogenesi e lo sviluppo, la percezione della luce e la risposta a stress (Stracke, 2001).

Alcuni membri delle famiglie ERF, bZIP, WRKY e MYB sono gli attivatori trascrizionali maggiormente caratterizzati come responsabili delle risposte di difesa della pianta. Mentre alcuni di questi attivatori sono specificatamente coinvolti nella regolazione dei meccanismi di difesa altri agiscono coordinando i processi di sviluppo della pianta (Hegedus et al., 2003).

La proteina ARR2 è un attivatore della trascrizione stimolato da citochine che è in grado di legare i promotori di geni responsabili della difesa da attacchi di patogeni (Choi et al., 2010).

Nel nostro caso non c'è una interazione pianta-parassita ma fattori di trascrizione di questo tipo potrebbero intervenire durante l'instaurazione della simbiosi che rimane comunque un meccanismo di interazione ospite-pianta.

L'unica sequenza che cataloghiamo come UA è la calmodulina. Le piante hanno sviluppato un elaborato sistema per la percezione dell'ambiente endogeno ed esogeno in modo da garantirsi una risposta fisiologica adeguata e spesso sfruttano il calcio come secondo messaggero. La calmodulina è una proteina che nella sua forma legata al Ca^{2+} regola diversi processi cellulari. In particolare la calmodulina può essere localizzata nel nucleo dove è in grado di legare alcune proteine, tra cui fattori di trascrizione, e di avere un ruolo, attraverso la regolazione genica, nella risposta cellulare a stimoli di carattere ambientale o di sviluppo. Ad esempio un fattore di

trascrizione WRKY coinvolto nella difesa della pianta da attacchi patogeni interagisce con la calmodulina in modo Ca2+-dipendente (Kim et al., 2009).
Tra le 18 sequenze uniche appartenenti alla categoria PUAs ritroviamo 12 proteine a funzione sconosciuta, due transferasi, una chinasi.

Analisi funzionale dei TF maggiormente espressi in ectomicorriza e corpo fruttifero

Lo *screening* mediante TAT può essere sfruttato non solo per l'analisi di genoteche ma anche per lo studio di una singola proteina di cui si è interessati sapere se è un attivatore della trascrizione.
Per validare alcuni tra i TF che nelle analisi array e RNA-seq risultavano essere espressi in maniera preferenziale in corpo fruttifero ed ectomicorriza rispetto al micelio (tabella 3 in appendice) si è deciso di clonare queste sequenze nel vettore pDEST32 in modo da valutare il fenotipo di attivazione dei tre geni *reporter*. Le amplificazioni dei TF sono state effettuate tramite *touch-down* PCR partendo dalle genetoche in cui erano espresse maggiormente. I primer utilizzati contenevano siti di restrizione CpoI, in modo da avere clonaggi direzionali.
I TF sono stati analizzati mediante un *drop assay* su terreno SD-Leu-His addizionato di 3-AT alle concentrazioni 10, 25, 50, 100 e 200 mM, su SD-Leu-Ura e con il test *LacZ*.
Per questa analisi mirata sono stati scelti 3 TF appartenenti alla classe pTF: TmelAcuK che è il regolatore della gluconeogenesi, TmelFF7 che è coinvolto nello sviluppo sessuale e TmelCbf1 che è un regolatore del metabolismo dello zolfo. Sono stati selezionati anche 5 CH-TF e 6 fattori di trascrizione unici di *Tuber* tra cui XlnR2. Quest'ultimo è molto interessante poiché potrebbe rappresentare un attivatore degli enzimi che degradano la parete cellulare vegetale e perché, essendo sovraespresso nella fase simbiotica, potrebbe rappresentare una proteina unica di *Tuber* coinvolta nell'instaurazione della simbiosi con la pianta. Rappresenterebbe una peculiarità di questo fungo anche se il gene deve essere ancora caratterizzato.

Fig. 6: *Spot assay* su terreno SD – Leu –His 3AT 0, 10

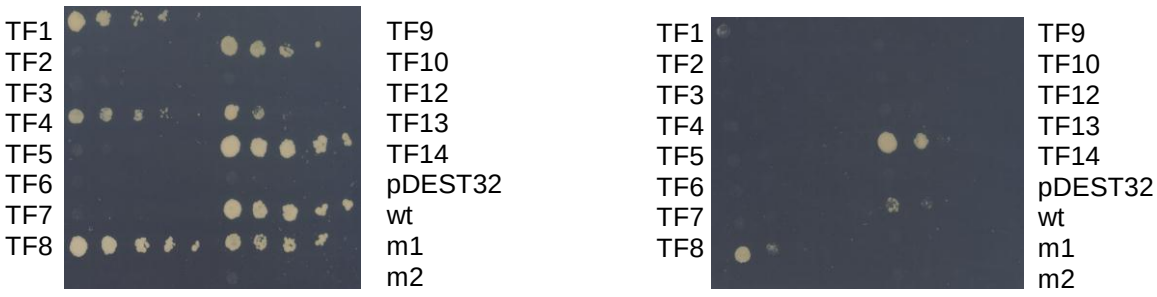


Fig. 7: *Spot assay* su terreno SD – Leu –His 3AT 50 e 200mM

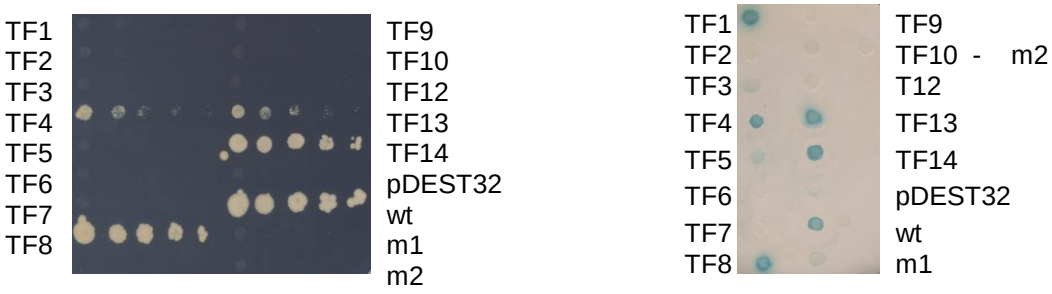


Fig. 8: *Spot assay* su terreno SD –Ura e saggio LacZ

Gene model	Name	DBD	espr	Nome	LacZ	Ura	3AT	TAT
GSTUMT00000111001	TmelACUK	C6 *zinc cluster*	FB	TF1	blu	no	100mM	FORTE
GSTUMT00000279001	HP-DBD	C2H2	ECM	TF2	bianco	no	no	non attiva
GSTUMT00000814001	TmelCBF1	bHLH	FB	TF3	bianco	no	no	non attiva
GSTUMT00004451001	CH-TF	Homeobox	ECM	TF4	blu	poco	50mM	MEDIO
GSTUMT00005836001	HP-DBD	C6 *zinc cluster*	ECM	TF5	azzurro	no	25mM	DEBOLE
GSTUMT00008124001	TmelFF7	C6 *zinc cluster*	ECM	TF6	bianco	no	no	non attiva
GSTUMT00009188001	CH-TF	C6 *zinc cluster*	FB	TF7	bianco	no	10mM	non attiva
GSTUMT00009814001	HP-DBD	C6 *zinc cluster*	FB	TF8	blu	cresce	200mM	FORTE
GSTUMT00011475001	HP-DBD	Myb-like factors	ECM	TF9	bianco	no	no	non attiva
GSTUMT00012451001	TmelXLNR-2	C6 *zinc cluster*	ECM	TF10	bianco	no	50mM	DEBOLE
GSTUMT00004341001	CH-TF	C6 *zinc cluster*	ECM	TF12	bianco	no	no	non attiva
GSTUMT00004776001	HP-DBD	Forkhead	FB	TF13	blu	poco	50mM	MEDIO
GSTUMT00006234001	CH-TF	CP2	ECM	TF14	blu	cresce	200mM	FORTE

Tabella 6: Risultato del TAT mirato

I risultati dello *screening* mostrano come tre TF (TmelAcuK , un CH-TF e un HP-DBD) siano attivatori forti mentre un CH-TF e una HP-DBD siano attivatori medi, due HP-DBD tra cui anche XlnR2 sono attivatori deboli mentre tutti gli altri non sono in grado di attivare l'espressione genica.

La mancata espressione genica può essere dovuta: a una modificazione post-traduzionale che la proteina deve subire per poter essere attiva e svolgere la sua funzione di attivatore, alla mancanza nel sistema eterologo di un interattore del fattore di trascrizione in esame oppure alla fusione con il DBD di Gal4 che rende in qualche modo la proteina inattiva.

Grazie a questo esperimento di TAT mirato i fattori di trascrizione validati funzionalmente aumentano a 44 su 201 totali (22%).

3. *Nuclear transportation trap* (NTT)

Nell'analisi TAT si individuano proteine che non sono note essere coinvolte nel processo di trascrizione; questi cloni possono rappresentare dei falsi positivi metodologici oppure proteine per cui si è evidenziata una funzione diversa rispetto a quella attribuita ad essa. In questo caso si potrebbero evidenziare nuovi esempi di proteine *moonlighting* che mantengono la funzione originaria pur essendo in grado di svolgere altri ruoli.

È però possibile ottenere anche dei cloni falsi positivi di tipo metodologico che sono dovuti al fatto che il dominio di DNA di Gal4, a cui vengono fusi i cDNA della genoteca, possiede un segnale di localizzazione nucleare proprio che forza l'espressione della proteina di fusione nel nucleo. Queste proteine in realtà non riuscirebbero naturalmente ad entrare nel nucleo perché mancano di un proprio NLS. Diventa fondamentale riuscire a capire se le proteine che emergono dallo *screening* TAT fanno parte dei falsi positivi o se possiamo ipotizzare di attribuire a questi cloni una nuova funzione non ancora scoperta.

Per questo motivo il primo passo per l'analisi di queste sequenze prevede l'utilizzo di un saggio di *Nuclear transportation trap* (NTT) (Marshall et al., 2007) che consente la validazione della presenza di un segnale di localizzazione nucleare all'interno della sequenza in esame. Come precedentemente descritto nella sezione materiali e metodi anche questo saggio viene condotto in lievito e utilizza un vettore centromerico che permette l'espressione di una tripla proteina di fusione tra la sequenza batterica di *LexA* modificata (ΔNLS), il dominio di attivazione Gal4AD e il cDNA della proteina da saggiare clonata in frame a valle di Gal4AD. Solo se la proteina in esame possiede un

NLS autonomo riesce a far traslocare la proteina di fusione nel nucleo e ad attivare l'espressione dei geni *reporter HIS3 e LacZ;* i quali posseggono a monte il sito operatore LexA riconosciuto dal *DNA binding domain* di LexA.

Alcune delle sequenze appartenenti alle categorie UAs e PUAs sono state saggiate con questa metodologia poiché si desidera sia verificare la localizzazione nucleare delle proteine di *Tuber* che possiedono un omologo con funzione nucleare documentata sia predire una localizzazione nucleare per proteine per cui non era ancora nota. Inoltre è necessario utilizzare anche un controllo positivo testando un fattore di trascrizione che permetta di verificare che il saggio sia condotto nelle condizioni ottimali. Sono state saggiate le sequenze PUAs che codificano per sei enzimi del metabolismo perché potrebbero rappresentare buoni candidati per l'isolamento di nuove *moonlighting protein.* In lettartuta troviamo infatti un buon numero di esempi di enzimi multifunzionali (come evidenziato nell'introduzione). Sono state saggiate due chinasi: Maf1 appartenente alla categoria UAs che funge sia da controllo sia per verificare che anche la proteina di *Tuber* abbia localizzazione nucleare e Ste50 che è una chinasi coinvolta nella trasduzione del segnale durante il processo di *mating* ed è nota avere azione nel citoplasma. Per la proteina di *Tuber* GSTUMT00005390001 dobbiamo sottolineare che eseguendo analisi BLAST la proteina risulta omologa alla MAP chinasi Maf1 mentre il *best reciprocal hit* in *S. cerevisiae* risulta essere la proteina Maf1 che è un regolatore negativo della RNA polimerasi III. Analizzando in modo approfondito gli allineamenti si deduce che la proteina di *Tuber* presenta maggiori omologie con le MAP chinasi. Per questo motivo viene analizzata come se avesse tale funzione all'interno della cellula.

Inoltre tra le sequenze appartenenti alle PUAs sono state analizzate anche la calcineurina e la perossina 20 mentre per le UAs sono state saggiate l'actina e la sequenza omologa alla proteina Lvt1.

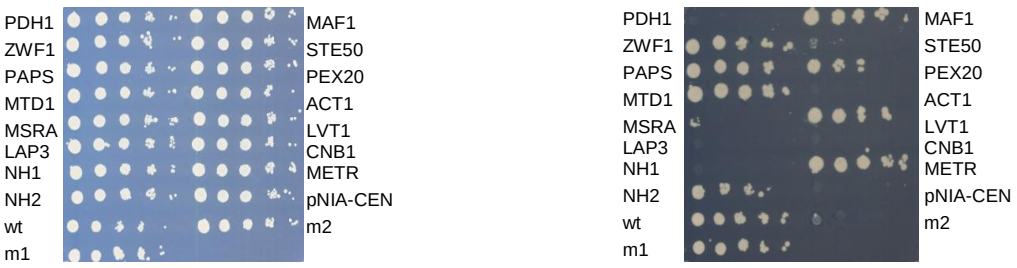


Fig. 9: *Screening* dei geni *reporter* effettuati mediante *drop assay* su SD –Leu, SD –Leu –His 10mM

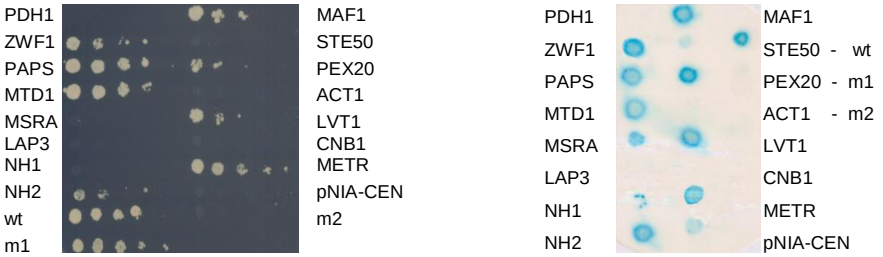


Fig. 10: *Screening* dei geni *reporter* effettuati mediante *drop assay* su SD –Leu –His 50mM e saggio LacZ.

gene model	classe	Sc homolog	blast-P	e-value	NLS PSORT	3AT	LacZ
GSTUMT00008609001	**pTFs**	**MetR**	**bZIP transcription factor (MetR)**	**3E-31**	**Y**	**50mM**	**blu**
GSTUMT00001950001	**UAs**	**Ltv1**	**low-temperature viability protein ltv1**	**9E-73**	**Y**	**50mM**	**blu**
						non	
GSTUMT00006162001	UAs	Act1	actin	6E-20	N	cresce	bianco
			peptide methionine sulfoxide reductase			non	
GSTUMT00004791001	UAs	Mxr1	msrB/msrA	8E-74	N	cresce	blu
						non	
GSTUMT00003999001	PUAs	Lap3	bleomycin hydrolase, putative	6E-156	Y	cresce	bianco
						non	
GSTUMT00004644001	PUAs	Pdh1	2-methylcitrate dehydratase	0	N	cresce	bianco
GSTUMT00008696001	**PUAs**	**Zwf1**	**glucose-6-phosphate 1-dehydrogenase**	**0**	**N**	**50mM**	**blu**
			3-phosphoadenosine-5-phosphosulfate				
GSTUMT00002663001	**PUAs**	**Met16**	**reductase**	**1E-111**	**N**	**50mM**	**blu**
			5,10-methylenetetrahydrafolate				
GSTUMT00003049001	**PUAs**	**Mtd1**	**dehydrogenase**	**3E-99**	**N**	**50mM**	**blu**
GSTUMT00002382001	**PUAs**	**Pex20**	**peroxin 20**	**5E-25**	**N**	**50mM**	**azzurro**
GSTUMT00005390001	**UAs**	**Maf1**	**mitogen-activated protein kinase MAF1**	**8E-63**	**Y**	**50mM**	**blu**
			MAPKKK cascade protein kinase			non	
GSTUMT00010173001	PUAs	Ste50	regulator Ste50	1E-83	Y	cresce	azzurro
						non	
GSTUMT00005170001	PUAs	Cnb1	calcineurin regulatory subunit B	2E-83	N	cresce	bianco
						non	
GSTUMT00003548001	PUAs				Y	cresce	bianco
GSTUMT00007946001	**PUAs**				**N**	**50mM**	**blu**

Tabella 7: cloni selezionati per l'analisi NTT, in grassetto sono le sequenze risultanti positive al saggio.

Come è riassunto in tabella sono risultate avere un segnale di localizzazione 8 proteine sulle 15 saggiate. I risultati ottenuti sperimentalmente non in tutti i casi coincidono con le predizioni del programma bioinformatico PSORT. Questo è però consistente con quanto descritto nell'introduzione per cui esistono dei segnali di localizzazione nucleare non canonici che non vengono predetti *in silico* mentre ci sono proteine che hanno un NLS predetto che non viene riconfermato dall'analisi NTT. Quest'ultimo caso può essere dovuto dalla mancanza di modificazioni che la proteina deve subire per poter esporre il segnale di localizzazione, mancanza di interattori, interazioni con proteine citoplasmatiche che la ancorano e ne bloccano la traslocazione nel nucleo o più semplicemente viene predetto un NLS che nella cellula non è funzionale.
Una dimostrazione dell'attendibilità dei risultati è fornita dal fattore di trascrizione e dalla MAP chinasi che attivano entrambi i geni *reporter*. Anche la UAs TmelLtv1 e una PUAs che codifica per una proteina unica di *Tuber* a funzione sconosciuta possiedono un NLS autonomo. Come si nota dalla tabella tra le due proteine uniche del fungo saggiate risulta positiva all'NTT quella per cui con i metodi bioinformatici non era predetto un NLS mentre quella negativa all'analisi possedeva un NLS predetto.
Si nota anche che il 50% degli enzimi saggiati possiede un segnale di localizzazione autonomo. Questi enzimi potrebbero essere interessanti per ulteriori analisi che ne confermino la funzione di attivatori della trascrizione ad esempio provando a saggiare la capacità di legame al DNA di queste proteine.
Tra le proteine che non possiedono un NLS ritroviamo 3 enzimi, la MAP chinasi Ste50, la calcineurina e una proteina unica di *Tuber* mentre tra le UAs l'actina sembra non possedere un NLS.
Il risultato riguardante l'actina era inatteso poiché in letteratura ci sono numerosi esempi di ruoli di questa proteina nel nucleo e in particolare nella trascrizione. Questo conduce a pensare a due ipotesi; la prima che la sequenza di *Tuber* non riesca realmente ad entrare nel nucleo mentre la seconda che l'actina, essendo espressa ad un basso numero di copie, si leghi preferenzialmente alle proteine del citoplasma con cui potrebbe avere maggiore affinità e che non riesca a traslocare nel nucleo.
La calcineurina invece può essere catalogata come falso positivo dovuto alla sua espressione forzata nel nucleo. La calcineurina è nota legare fattori di trascrizione nel citoplasma e probabilmente risulta positiva al TAT perché una volta forzata nel nucleo riesce a interagisce con questi fattori di trascrizione e risulta in grado di attivare la trascrizione dei geni *reporter*.

4. Analisi della fosfoadenosina-fosfosolfato reduttasi

Dato il crescente interesse per le proteine multifunzionali e visto i numerosi studi che evidenziano come enzimi citoplasmatici possono avere un ruolo anche in altri processi cellulari è importante approfondire lo studio delle proteine non convenzionali che emergono dalle analisi TAT e NTT per comprendere il loro reale coinvolgimento nella attivazione della trascrizione genica.

In particolare è stata studiata la fosfoadenosina-fosfosolfato reduttasi (PAPS reduttasi) per capire se le peculiarità di questa proteina, che ha inaspettatamente una localizzazione nucleare ed è coinvolta nell'attivazione trascrizionale, sono proprie solo della proteina di *Tuber* o se sono comuni anche agli omologhi di *Neurospora crassa* e *Saccharomyces cerevisiae.*

La PAPS reduttasi è un enzima coinvolto nell'assimilazione del solfato. Il solfato è la maggior fonte di zolfo per molti organismi come batteri, funghi e piante che sono in grado di ridurre il solfato in solfito e successivamente di ridurre il solfito a solfuro. Per poter ridurre il solfato a solfito è però necessario attivare il solfato legandolo all'ATP per generare l' APS (adenil-fosfosolfato). L'APS viene successivamente fosforilata a PAPS che viene ridotta in solfito e PAP dalla PAPS reduttasi. Il solfito viene poi ridotto in solfuro ed è utilizzato dalla cellula per la biosintesi della cisteina e della metionina (Yu et al., 2008).

Per l'analisi della PAPS reduttasi di *Tuber* il primo approccio, dopo il *nuclear transportation trap,* è stato quello di verificare che la proteina possedesse l'attività enzimatica tramite un saggio di complementazione in lievito.

Per questo scopo sono state clonate nel vettore di espressione di lievito pYX212 le sequenze della PAPS di *T. melanosporum* e di *S. cerevisiae* per effettuare il saggio di complementazione nel ceppo di lievito CC362-2A.

In *S. cerevisiae* la delezione del gene che codifica per la PAPS reduttasi causa auxotrofia per la metionina; quindi utilizzando questo ceppo è possibile verificare la complementazione attraverso il ripristino della capacità di crescita in assenza di metionina.

Nell'esperimento di complementazione sono state testate anche le PAPS reduttasi di *Tuber* e lievito fuse alla GFP poiché è stato condotto anche un esperimento di localizzazione mediante fluorescenza e si voleva verificare che la proteina di fusione non creasse dei problemi alla funzionalità della PAPS reduttasi.

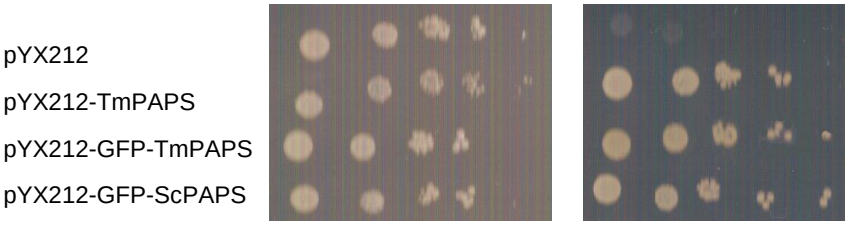


Fig. 11 : prova di complementazione su terreno SD –Ura +Met e SD –Ura –Met

Come emerge in figura la PAPS reduttasi di *T. melanosporum* complementa e quindi possiede l'attività enzimatica. Inoltre la fusione alla GFP non danneggia la funzionalità delle due proteine saggiate.

Si sono poi eseguiti esperimenti di TAT mirato e di NTT per verificare se anche le proteine di *S.cerevisiae* e *N.crassa* hanno localizzazione nucleare e attivavano la trascrizione.

Per prima cosa si è eseguito uno *screening* di *Transcriptional Activator Trap* per valutare la capacità delle due PAPS reduttasi di *N.crassa* e lievito di attivare la trascrizione. I cDNA codificanti per queste proteine sono stati clonati nel vettore pDEST32, trasformati nel lievito Mav103 e saggiati mediante *drop assay.*

Dai risultati è emerso che solo la PAPS reduttasi di *Tuber* è in grado di attivare l'espressione dei geni *reporter.* Quindi la funzione di attivatore è una sua peculiarità.

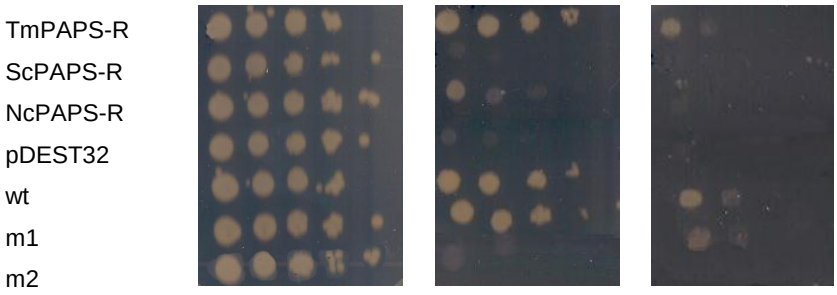


Fig. 12: Drop assay su SD –Leu, SD –Leu –His +3AT 10mM e SD –Leu –His +3AT 100mM; TmPAPS-R: PAPS reduttasi di *T. melanosporum*, ScPAPS-R: reduttasi di *S.cerevisiae*, NcPAPS-R: reduttasi di *N.crassa*, pDEST32: vettore vuoto, controlli wt, m1, m2.

Ci si chiede se anche la localizzazione nucleare sia una caratteristica della proteina del fungo e per questo sono stati condotti due esperimenti: l'NTT per verificare la presenza di un NLS all'interno delle proteine e un saggio di localizzazione mediante fusione con la GFP.

Il *nuclear transportation trap* è stato condotto solo sulla PAPS reduttasi di *Tuber* e *Neurospora* poiché non si è ottenuto il clonaggio di quella di lievito.

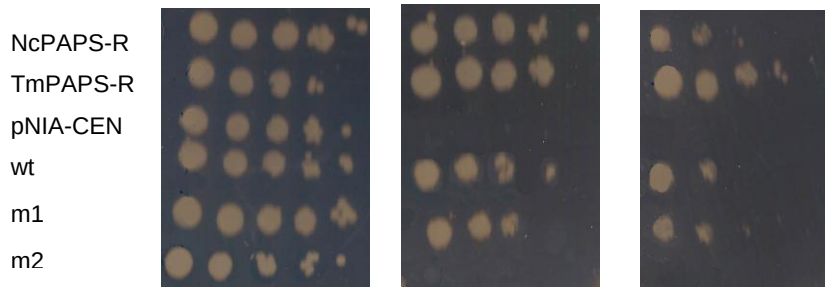


Fig.13: risultato *drop assay* NTT delle PAPS reduttasi di *N.crassa* (NcPAPS-R) e di *T.melanosporum* (TmPAPS-R) su terreno SD –Leu, SD –Leu –His 10 e 50mM

Da questo saggio è possibile concludere che anche la PAPS reduttasi di *Neurospora* possiede un segnale di localizzazione nucleare anche se non è coinvolta nell'attivazione della trascrizione.

Per determinare, in cellule di lievito, la localizzazione della proteina mediante fusione alla GFP si è trasformato il ceppo *wild type* W303-1A con i vettori di espressione pYX212 al cui interno erano state clonate le proteine di fusione GFP-PAPS reduttasi di *T.melanosporum*, *N.crassa* e *S.cerevisiae*. Le cellule cresciute *over-night* in terreno minimo sono state analizzate al microscopio a fluorescenza. Dalle immagini ottenute non è chiara la localizzazione cellulare in quanto è necessario ricordare che la PAPS reduttasi è sia citoplasmatica che nucleare. Con la sola fluorescenza della GFP il segnale derivante dalla proteina citoplasmatica e da quella nucleare si sovrappongono. Da questo esperimento si può solo dedurre la localizzazione nucleare della PAPS reduttasi di *Tuber* poiché sembrano emettere fluorescenza anche i nuclei. Per le altre due proteine invece è più difficile l'interpretazione dell'immagine ottenuta al microscopio.

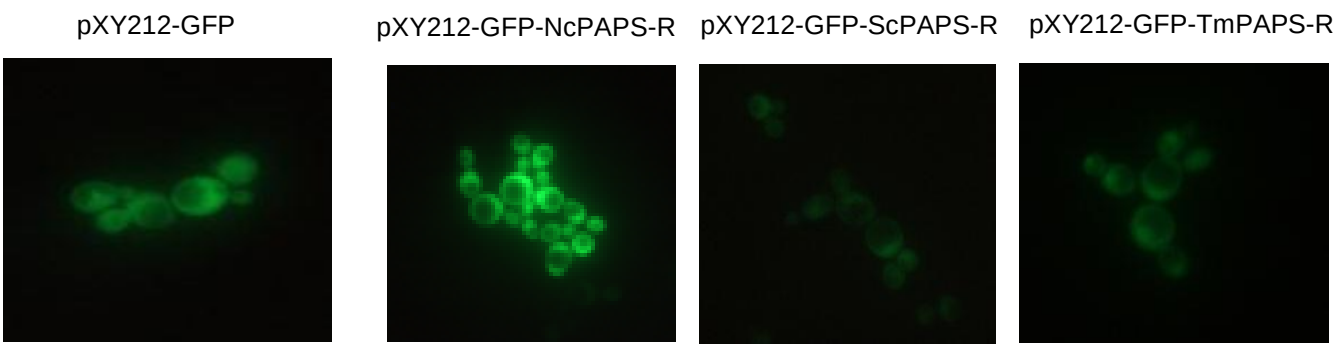


Fig. 14: immagini al microscopio a fluorescenza

Si nota analizzando l'allineamento delle tre sequenze che la maggior parte della proteina risulta essere conservata mentre le uniche differenze evidenti sono presenti alle estremità. Infatti le proteine di *Tuber* e *Neurospora* possiedono una regione al C-terminale e all'N-terminale che non sono presenti nella sequenza di lievito.

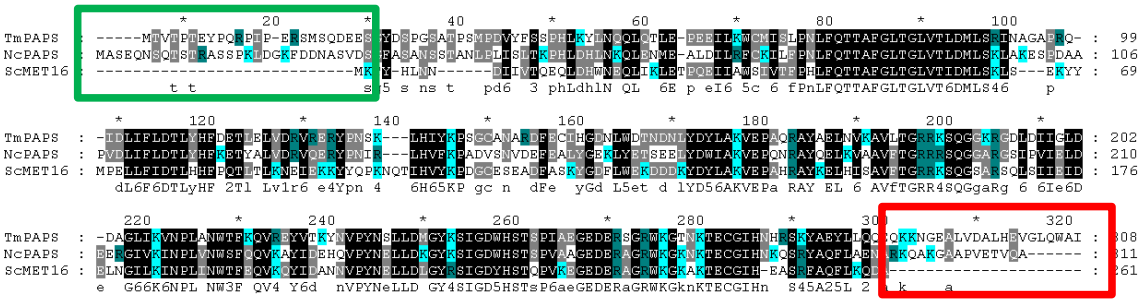


Fig. 15: allineamento delle PAPS reduttasi di *T.melanosporum* (TmPAPS), *N. crassa* (NcPAPS) e di *S.cerevisiae* (ScMet16)

Potrebbe essere una di queste due regioni che conferisce alla PAPS reduttasi la capacità di attivare la trascrizione. Per questo motivo sono state analizzate attraverso TAT e NTT le proteine delete in queste due regioni per vedere se la proteina mutata perdeva la capacità di attivare la trascrizione. Sono stati creati quindi due costrutti: pDEST32-ΔNPAPS-R in cui manca la regione della proteina N-terminale (nella figura 15 evidenziata nel riquadro verde) e pDEST32-ΔCPAPS-R deleta al C-terminale (la regione eliminata è quella nel riquadro rosso della figura 15).

Per quanto riguarda i risultati del TAT sulle delezioni si è dimostrato che, se deleta della regione C-terminale, la proteina perde la capacità di attivazione della trascrizione. La funzione di attivatore della trascrizione è però mantenuta nella proteina deleta all'N-terminale anche se l'attivazione del gene *reporter* risulta essere leggermente più debole in quanto la proteina intera a 100mM mostra maggior crescita delle diverse diluizioni.

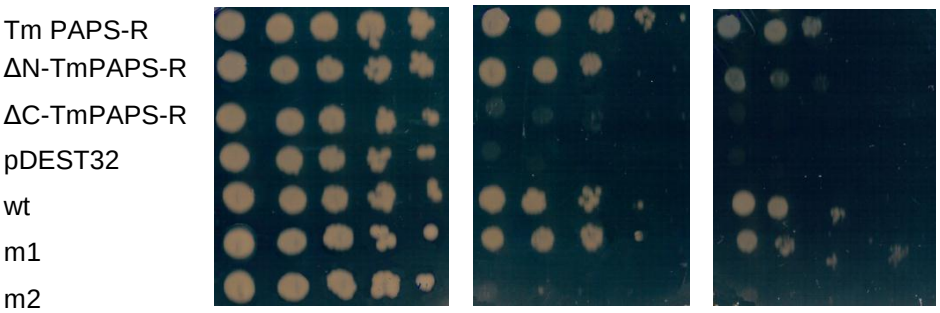


Fig. 16: Risultati TAT per le proteine delete, *drop assay* su terreno SD –Leu, SD –Leu –His 25 e 100mM

Per quanto riguarda invece le analisi NTT entrambe le delezioni mantengono la capacità di entrare nel nucleo. Questo significa che la regione contenente l'NLS si trova all'interno della sequenza proteica e non nelle regioni terminali.

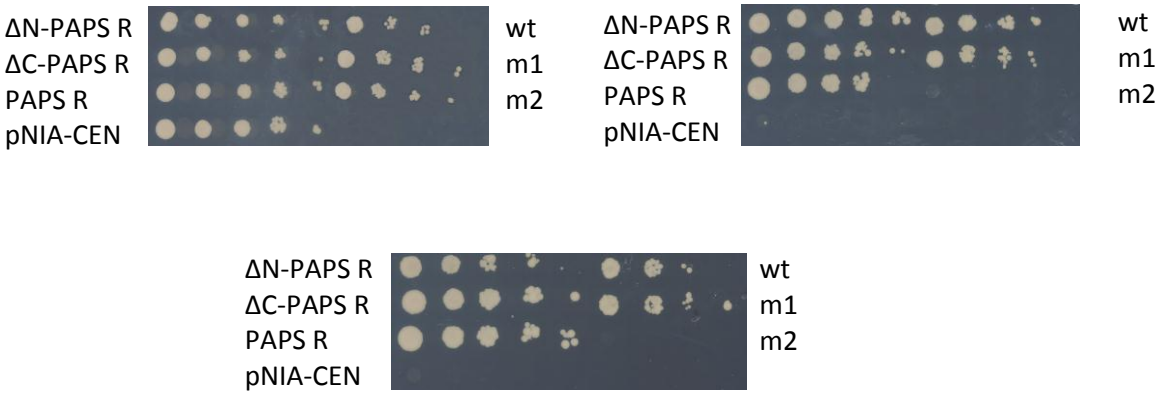


Fig. 17: *Nuclear transportation trap* della PAPS reduttasi e delle forme delete al C-terminale e all'N-terminale, *Dropo assay* su terreno SD –Leu, SD –Leu –His 10 e 50mM

La proteina deleta al C-terminale è stata utilizzata per esperimenti di complementazione nel ceppo CC362-2A. Si è verificato il mantenimento dell'attività enzimatica poiché il ceppo trasformato con questo costrutto riusciva a crescere su terreno selettivo in assenza di metionina.

Questo risultato è fondamentale in vista di una possibile attribuzione di questa proteina alla categoria delle *moonlighting* in quanto, come sottolineato nell'introduzione, un requisito delle proteine multifunzionali è che le diverse funzioni siano assolutamente indipendenti una dall'altra e che la mutazione che causa la perdita di una di queste non influisca sulle altre attività.

Dimostrato che è la regione C-terminale della PAPS reduttasi responsabile dell'attivazione dell'espressione genica ma non della sua attività enzimatica si è testata la capacità della sola regione C-terminale di attivare l'espressione genica. A questo scopo è stato clonato il frammento in frame con il dominio di legame al DNA di Gal4 all'interno del vettore pDEST32 e sono stati eseguiti *drop assay* per determinare l'attivazione dei geni *reporter*. Il frammento non è, da solo, in grado di attivare la trascrizione; questo può essere dovuto ad una non corretta conformazione del dominio poiché manca la maggior parte della sequenza proteica oppure il frammento non rappresenta l'intero dominio di attivazione.

Si è deciso quindi di creare una proteina di fusione tra la PAPS di *Neurospora*, che non è in grado di attivare l'espressione, e la PAPS reduttasi di *Tuber.* La regione C-terminale della PAPS di *Neurospora* è stata sostituita con la regione C-terminale della proteina di *Tuber* per verificare se questo dominio fosse capace di conferire alla proteina di *Neurospora* il ruolo di attivatore. È stata scelta per questo esperimento la PAPS di *Neurospora* poiché per tale proteina, a differenza di quella di lievito, è stata dimostrata la localizzazione nucleare. In questo modo i risultati ottenuti dalle analisi TAT sulla proteina di fusione non rischiavano di poter essere interpretati come falsi positivi metodologici.

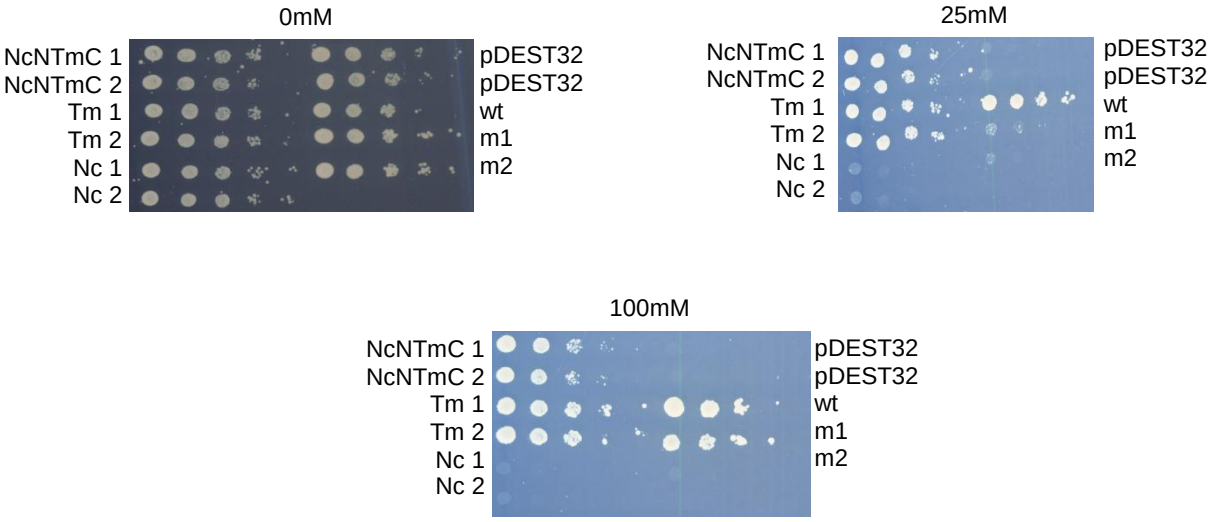


Fig. 18: risultato *screening* gene *reporter HIS3*

La proteina di fusione, a differenza della PAPS reduttasi di *Tuber*, è in grado di attivare l'espressione del solo gene *reporter HIS3* in modo forte. Come si vede nella figura il ceppo cresce sino a 100mM, mentre non attiva i geni *URA3 e LacZ.* Emerge in modo chiaro che la proteina di fusione, a differenza di quella di *Neurospora*, riesce a conferire attivazione forte di almeno un gene *reporter*. Per tutti questi motivi si può dedurre che il frammento C-terminale possa essere il responsabile dell'attivazione genica anche se in modo parziale quando fuso ad altre proteine.

Concludendo possiamo affermare che la fosfoadenosina-fosfosolfato reduttasi di *T.melanosporum* ha sia il ruolo enzimatico atteso sia funzione di attivatore e che le due funzioni sembrano essere indipendenti tra loro. La perdita della capacità di attivare la trascrizione non influenza il ruolo enzimatico.

5. *Bacterial one-hybrid*: analisi dei motivi di legame al DNA di fattori di trascrizione di *T. melanosporum*

Ora che sono stati catalogati i fattori di trascrizione codificati dal genoma di *Tuber melanosporum* e che sono state isolate interessanti proteine non convenzionali candidate ad avere un ruolo nella regolazione della trascrizione è interessante studiare quali sono le sequenze regolatrici riconosciute da queste proteine. Per i fattori di trascrizione che contengono un DBD noto e che hanno caratteristiche interessanti come TmelXlnR2 e come le tre copie di TmelNirA è importante capire quali sequenze di DNA legano per determinare possibili peculiarità che contraddistinguono questo fungo. Nel caso di TmelXlnR2, un fattore di trascrizione unico di *Tuber*, il poter evidenziare i motivi di legame e di conseguenza anche i geni regolati da tale proteina permetterebbe di capire se è coinvolto nella regolazione degli enzimi che degradano la parete, come ipotizzato, e soprattutto di quali enzimi si tratta.
Nel caso delle tre copie del TF TmelNirA si vuole determinare se legano sequenze diverse per capire il motivo di questa duplicazione genica all'interno del genoma dato che gli altri funghi presentano una singola copia del gene.
La tecnica del *bacterial one-hybrid* prevede l'espressione del DBD o dell'intera proteina all'interno di una cellula batterica contenente una vettore che porta una sequenza casuale a monte di due geni *reporter*, se il TF lega una delle sequenze presenti nella *library* si otterrà l'espressione del gene *reporter HIS3* che ne consentirà la crescita su terreno selettivo contenente 3AT. La *library* prima di essere utilizzata deve essere ripulita in modo da eliminare gli auto-attivatori cioè tutte quelle proteine presenti nella cellula batterica che sono in grado di attivare l'espressione dei geni *reporter* in assenza dell'esca. Per questa selezione si sfrutta il gene *reporter URA3* che quando è espresso metabolizza in un composto tossico il 5-FOA presente nel terreno selettivo. Questa tecnica prevede l'isolamento di motivi di legame al DNA in vivo e con una metodologia rapida e facilmente applicabile in tutti i laboratori.
Nonostante ciò per le proteine di *Tuber* saggiate sono sorti numerosi problemi che non hanno permesso un efficace utilizzo del metodo.

Le proteine che sono state saggiate sono:

ESCHE	**DBD**	**DESCRIZIONE**
Zif268 pB1H2ω2	C2H2	controllo positivo
mutOdd pB1H2ω2	C2H2	controllo negativo
NirA2-DBD pB1H2ω5	C6 zn cluster	nitrate regulatory protein
NirA1-DBD pB1H2ω5	C6 zn cluster	nitrate regulatory protein
NirA1-FL pB1H2ω5	C6 zn cluster	controllo interno per verificare se è possibile clonare anche la sequenza FL
TF10-FL pB1H2ω5	C6 zn cluster	TmelXLNR-2, ECM up
TF5-DBD pB1H2ω5	C6 zn cluster	HP_GSTUMT-05836001, ECM up
TF13-DBD pB1H2ω5	fork head	HP_GSTUMT-04776001, FB up
TF8-DBD pB1H2ω5	C6 zn cluster	HP_GSTUMT-09814001, FB up
PAPS pB1H2ω5	-	forte attivatore trascrizione privo di un DBD
ScRds1-DBD pB1H2ω5	C6 zn cluster	TF di *S.cerevisiae* utilizzato come controllo interno per i C6
ScMet32-FL pB1H2ω5	C2H2	TF di *S.cerevisiae* utilizzato come controllo interno per i C2
TmMet32-DBD pB1H2ω5	C2H2	Activator of sulfur metabolism, utilizzato come controllo interno

Tabella 8: Proteine clonate nel vettore di espressione per essere saggiate attraverso esperimenti di B1H

Come raffigurato in tabella sono stati testati i due controlli Zif268pB1H2ω2 (positivo) e mutOddpB1H2ω2 (negativo) forniti con il materiale del B1H e due geni coinvolti nel metabolismo dell'azoto TmelNirA1 e TmelNirA2. TmelNirA3 non è stato clonato poiché la sua sequenza anche se ristretta al solo DBD possedeva al suo interno il sito di restrizione necessario per il clonaggio. Inoltre per quanto riguarda NirA1 si è deciso di clonare sia la sequenza completa che il DBD per verificare se con l'intera sequenza si verificavano problemi di espressione o se il saggio si poteva estendere senza troppe difficoltà anche a proteine con dimensioni maggiori rispetto ad un singolo dominio.
È stato analizzato anche il fattore di trascrizione unico del fungo TmelXlnR2. Per tale proteina si è deciso di clonare l'intero gene perché non è stata possibile una precisa predizione di un eventuale dominio di dimerizzazione; dato che la maggior parte dei

fattori C6 *zinc cluster* dimerizza l'analisi del DBD in assenza del dominio di dimerizzazione poteva risultare in un fallimento dello *screening.*

Sono stati scelti anche 3 TF tra quelli maggiormente espressi in corpo fruttifero (TF8 e TF13) e in ectomicorriza (TF5) che non hanno una funzione nota. L'identificazione dei motivi da loro riconosciuti e la successiva mappatura di questi motivi sul genoma consentirebbe l'attribuzione di una funzione a questi geni particolarmente interessanti poiché coinvolti nello sviluppo delle due fasi del ciclo vitale che caratterizzano il fungo.

Inoltre avendo condotto esperimenti mirati per determinare la funzione di attivatore trascrizionale della PAPS reduttasi si è deciso di saggiare anche tale proteina per cercare di capire se è in grado di legarsi al DNA. Avere un risultato positivo da questo *screening* consentirebbe un fondamentale passo in avanti per l'attribuzione di una funzione nella regolazione della trascrizione a questo enzima.

Come controlli ulteriori per verificare l'attendibilità del metodo sono stati studiati due fattori di trascrizione di lievito: Rds1 appartenente alla classe dei C6 *zinc finger* e Met32 un C2H2 per cui è già nota la sequenza di legame al DNA.

Come controllo interno di *Tuber* per quanto riguarda la classe dei C2H2 si è deciso di clonare il TF Met32 che è l'omologo di quello scelto in lievito.

Per prima cosa si è costruita la *library* come descritto nei materiali e metodi che è costituita da una sequenza casuale di 20 nucleotidi. A seguito della trasformazione di cellule batteriche si sono ottenuti $6*10^6$ trasformanti che hanno subito una espansione di 20 volte. La *library* così costruita rappresenterà solo una minima parte delle possibili sequenze di 20nt ma è opportuno considerare che il dominio di legame di un fattore di trascrizione è inferiore a 20nt, se per esempio consideriamo come lunghezza media di un motivo 8 paia di basi allora la *library* diventa sufficientemente rappresentativa perché il motivo casuale può trovarsi anche in diverse posizioni della sequenza.

Una volta ottenuta la *library* si è proceduto con la trasformazione nel ceppo di selezione USO e con l'eliminazione degli auto-attivatori attraverso selezione negativa su terreno selettivo contenente 5-FOA 2,5mM. Sono state sottoposte a selezione negativa un totale di 10^7 colonie che sono state suddivise in 40 piastre da 50ml. Le colonie cresciute dopo 13h su 5-FOA sono state raccolte utilizzando 10ml di H_2O per ognuno dei piastroni utilizzati. Dopo averne calcolato il titolo sono state piastrate $3*10^7$ colonie dividendole in 40 piastroni. In questo modo si è eseguito un secondo passaggio di eliminazione degli auto-attivatori. Questo secondo passaggio è stato reso indispensabile dal fatto che testando la *library* sottoposta a una sola ripulitura non si ottenevano differenze tra il numero di colonie cresciute dopo selezione positiva su 3AT

tra il controllo positivo Zif268-pB1H2ω2 e il controllo negativo mutOdd-pB1H2ω2 mentre secondo il protocollo è richiesta una differenza di almeno 10 volte.
Con diverse prove di ripulitura il miglior risultato ottenuto è stato un rapporto tra colonie cresciute del positivo e del negativo di 6 volte. Si sono riscontrati numerosi problemi di selezione su 5-FOA, con una sola pulitura sembra addirittura che non si riescano ad eliminare gli autoattivatori visto che non ci sono differenze di crescita tra i controlli.
A seguito della seconda ripulitura è stata saggiata la *library* eseguendo la selezione positiva su piastre contenenti 3AT solo con il controllo positivo Zif268-pB1H2ω2 e con il controllo negativo mutOdd-pB1H2ω2.
Le colonie positive alla selezione per l'esca Zif268 sono state analizzate e a seguito del sequenziamento è emerso che solo il 50% delle sequenze possedeva il motivo di legame di Zif268. Per questo si è deciso di implementare il protocollo utilizzando anche l'ulteriore procedura consigliata nel *Nature Protocol* (Meng and Wolfe, 2006) in caso di fallimento della selezione positiva. A seguito della selezione su 3AT le colonie sono state raccolte e si è proceduto con l'estrazione della miscela dei plasmidi esca e *library*. I vettori sono stati sottoposti a digestione con l'enzima di restrizione XmnI per eliminare l'esca. La miscela di digestione è stata usata per ritrasformare il ceppo USO ed è stata piastrata nuovamente su 5-FOA. Inoltre si è deciso di implementare ulteriormente il protocollo con un passaggio che prevede la raccolta dei trasformanti in piastre YT2X con kanamicina in formato a 96 e una volta cresciuti vengono replicati su NM addizionato di 3AT. Le colonie che crescono rappresentano auto-attivatori perché riescono a crescere su 3AT in assenza dell'esca mentre le colonie positive sono quelle che in tali condizioni non riescono a crescere sul terreno selettivo.
Con questo metodo la totalità delle sequenze analizzate per il controllo Zif268 possiede il motivo di legame al DNA di questo fattore di trascrizione. Oltre al controllo si è ottenuto un probabile motivo di legame solamente per due proteine tra tutte quelle saggiate: TmelNirA2-DBD e TF5-DBD.
Per prima cosa si è trasformato il ceppo USO contenente il TF con 100ng della *library* ripulita due volte e la trasformazione è stata divisa su piastroni NM 3AT 5 e 10mM piastrando 2ml per ogni piastra. In questo modo sono state vagliate $4,7*10^7$ colonie per il TF5-DBD, $4*10^8$ colonie per Zif268, $3,2*10^8$ colonie per mutOdd e 10^6 colonie per NirA2-DBD.

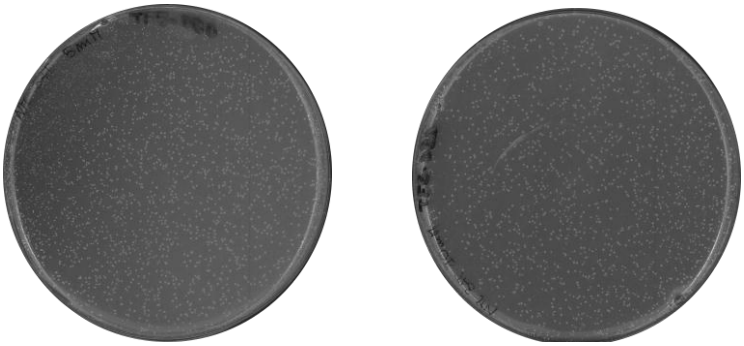


Fig. 19: selezione positiva del TF5-DBD + *library* su 3AT 5 e 10mM

Come si vede nella figura a parità di colonie piastrate su 3AT 10mM crescono meno colonie poiché aumentando la stringenza della selezione favoriamo via via legami più forti tra il fattore di trascrizione e il motivo della *library*.

Queste colonie vengono raccolte in 10ml di H_2O e processate con il protocollo di minipreparazione plasmidica, la miscela di plasmidi dell'esca e della library vengono sottoposti a digestione con l'enzima di restrizione XmnI che taglia solamente il vettore pB1H2. Il ceppo USO è stato trasformato con la miscela di digestione purificata e i trasformanti sono stati piastrati su terreno YM con 5-FOA 2,5mM. Le colonie cresciute sono raccolte in formato da 96 su terreno YT2X con kanamicina e successivamente sono state replicate su terreno 3AT 5 o 10mM in base alla piastra di selezione positiva. Abbiamo replicato 40 colonie contenenti la *library* per ogni trasformazione.

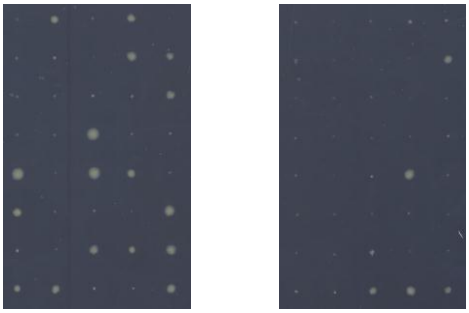


Fig.20: replica su NM 3AT 5 e 10mM delle 40 colonie della *library* positva a TF5-DBD

Come si vede nell'immagine per il TF5-DBD sono da considerarsi positive 23 colonie su 40 (57%) nelle condizioni di 3AT 5mM mentre con concentrazione più elevata (10mM) risultano positive 34 colonie (85%). Per gli altri fattori di trascrizione analizzati si ottengono: per Zif258 il 40% di positivi su 3AT 5mM e il 55% su 10mM, per NirA2-DBD otteniamo il 100% di positivi, mentre nel controllo negativo mutOdd il 92% è rappresentato da auto-attivatori.

Sono stati sequenziati 10 cloni della *library* positiva al controllo positivo Zif268 e tutti possedevano al loro interno il motivo noto essere riconosciuto da questa proteina.
Per quanto riguarda NirA2-DBD sono stati sequenziati 40 cloni e attraverso analisi con il programma bioinformatico MEME è emerso il seguente motivo:

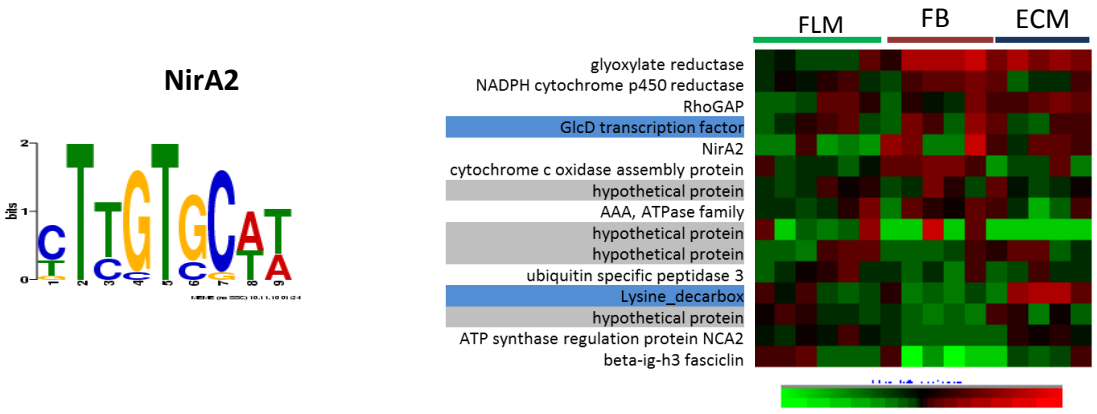


Fig. 21 : motivo di legame al DNA di TmelNirA2 e dati di espressione dei geni che presentano il motivo nel loro promotore

Si sono poi analizzate le sequenze promotrici del genoma per identificare in quali promotori si riscontra la presenza di tale motivo. Come atteso ritroviamo questo motivo a monte dei geni coinvolti nel metabolismo dell'azoto nitrito e nitrato reduttasi e nitrato permeasi ma sono presenti numerose copie di questo motivo anche nelle regioni genomiche a monte dei geni che codificano per la lisina decarbossilasi e per un fattore di trascrizione coinvolto nel metabolismo dell'azoto. Inoltre sembra che NirA2 regoli anche la sua espressione.
Per quanto riguarda invece il fattore di trascrizione TF5-DBD il motivo di legame al DNA è il seguente:

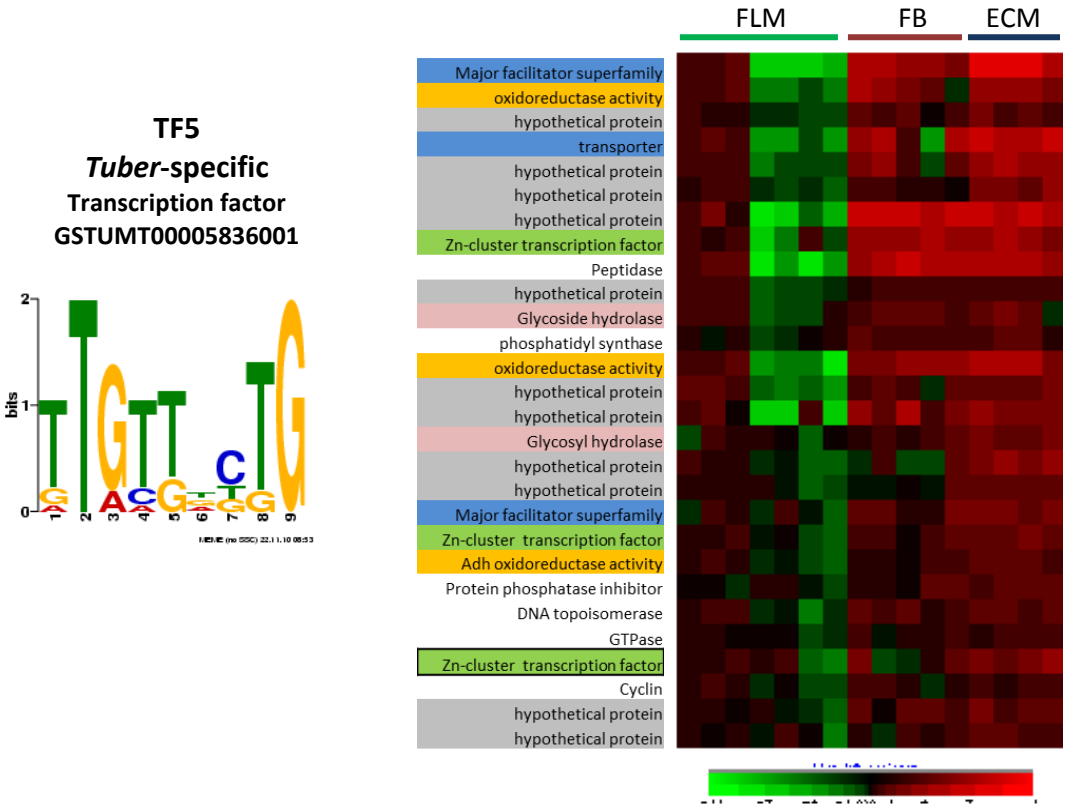


Fig.22 : motivo di legame al DNA di TF5 e gruppo di geni coregolati in cui si ritrova tale motivo

In questo caso l'analisi dei geni regolati da tale TF è più complessa perché essendo una proteina unica di *Tuber* e a funzione sconosciuta non abbiamo nessuna informazione preliminare che possa aiutare l'identificazione del ruolo del fattore di trascrizione. Una volta trovato il motivo con MEME e dopo l'analisi delle regioni promotrici del genoma si possono raggruppare i geni che potrebbero essere regolati da questo TF in base al loro profilo di espressione per determinare i geni che hanno lo stesso profilo del fattore di trascrizione. In questo caso si è interessati ai geni che sono sovraespressi nell'ectomicorriza. Un altro criterio è di isolare i promotori dove è presente il motivo in più copie e di escludere quelli in cui si ritrova un solo motivo.

Dall'analisi emerge che il motivo si trova nel promotore dello stesso TF5, ipotizzando che il fattore di trascrizione possa auto-regolarsi, nel promotore di altri due fattori di trascrizione C6 *zinc cluster* per cui non è nota la funzione, in due glicosil idrolasi, in tre proteine con attività di ossido-reduttasi, nei promotori di trasportatori, e in molti promotori di proteine a funzione sconosciuta.

Dai risultati ottenuti non è possibile dedurre un ruolo per questo fattore di trascrizione in quanto il suo motivo di legame al DNA non si trova in regioni a monte di geni con un ruolo ben determinato o di geni coinvolti nello stesso processo cellulare. Inoltre all'interno del genoma di *Tuber* sono molte le proteine non ancora caratterizzate funzionalmente. Risulta quindi alta la possibilità di isolare tali sequenze durante una analisi su scala genomica.

Per quanto riguarda le altre proteine analizzate non si sono ottenuti cloni positivi a seguito della selezione, tutte le colonie isolate su 5-FOA dopo digestione con XmnI erano caratterizzate da auto-attivatori.

Ci si è interrogati sui possibili motivi del fallimento del metodo e si sono ipotizzati diversi scenari. Il primo potrebbe essere un motivo di tossicità della proteina, per esempio le colonie che esprimevano il fattore di trascrizione NirA1 mostravano un fenotipo di crescita rallentato. Si è deciso allora di modificare il livello di espressione delle proteine NirA1-DBD, XlnR2, TmMet32-DBD e ScMet32 clonandole nel vettore pB1H2ω2 che le poneva sotto il controllo di un promotore debole. Questa modificazione non ha portato a nessun risultato positivo.

Il problema per alcuni fattori di trascrizione C6 *zinc finger* potrebbe essere causato da un legame del TF al vettore della *library* e non alla sequenza casuale poiché è noto che molti *zinc finger* riconoscono motivi CGG-$N_{N(3\text{-}23)}$-CCG e le regioni del vettore che fiancheggiano la sequenza casuale sono ricche in GC.

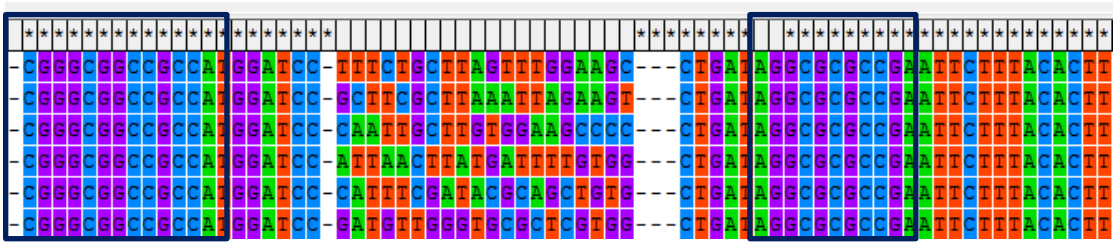


Fig. 23: rappresentazione della regione del vettore pB1H2 fiancheggiante l'inserto a sequenza casuale.

Il TF se si legasse alla regione a monte della sequenza casuale potrebbe essere troppo distante dalla box -35 del promotore debole posto a monte dei geni *reporter* per riuscire ad attivare l'espressione genica. Per questo motivo nella selezione si favorirebbe solo la crescita degli auto-attivatori. Oppure se il legame al vettore attivasse l'espressione dei geni *reporter* si otterrebbe l'isolamento di sequenze casuali che non permetterebbero l'identificazione di un motivo conservato. Per risolvere questo problema, almeno in parte, si potrebbe eliminare la regione ricca in GC posta a valle

delle sequenze casuali ricostruendo la *library* utilizzando il protocollo di *primer extension* e non quello utilizzato in questo caso di appaiamento di primer.
Un altro eventuale problema può essere rappresentato dalla dimerizzazione. Nel caso di omodimeri il problema viene eliminato andando a clonare anche il dominio di dimerizzazione mentre per eterodimeri si può condurre l'esperimento solo se è noto l'interattore. Inoltre per lo studio di dimeri Wolf e collaboratori utilizzano il sistema α poiché l'RNA polimerasi batterica possiede due di queste subunità al suo interno mentre per la subunità ω, presente in singola copia nell'RNA polimerasi, non ci sono esempi di studio di dimeri. Per ora, quindi, non sappiamo se il sistema ω possa essere idoneo a questo scopo.
Si è notato negli studi condotti da Wolf e collaboratori che un parametro importante è anche la distanza del motivo di legame dalla box -35 del promotore debole a monte dei geni *reporter*. Si nota come alcuni fattori di trascrizione leghino in modo preferenziale motivi situati il più vicino alla box -35 anche se ci sono esempi di proteine che spaziano lungo tutta la lunghezza della *library*. Inoltre la regione a sequenza casuale, per come è stata inserita nel vettore, è posta a 21bp dalla box -35 ma ricostruendola per *primer extension* la distanza si ridurrebbe a 7bp. Questa maggiore distanza della sequenza casuale potrebbe sfavorire l'identificazione dei motivi troppo distanti dal promotore. Per far fronte a questo problema, senza ripreparare la *library,* si è deciso di modificare il vettore di espressione del fattore di trascrizione in modo da aggiungere un linker tra il TF e la subunità ω in modo da avere un linker di collegamento tra i due peptidi più lungo. In questo modo si cerca di favorire il corretto posizionamento dell'RNA polimerasi sul promotore anche quando il motivo si trova nella regione della sequenza casuale più distante dalla box -35. Per ora questo costrutto è stato saggiato con il controllo positivo Zif268 per cui abbiamo verificato il corretto funzionamento della nuova proteina di fusione in cui il linker è più lungo di 13 aa. Per quanto riguarda il controllo positivo Zif268 non si è ottenuto un aumento dei cloni positivi e nemmeno una maggior distribuzione del motivo lungo tutta la sequenza ma è in programma, dato che abbiamo dimostrato che il linker non crea problemi di espressione, di saggiare TF per cui non si sono ottenuti risultati con questo nuovo costrutto.
Un'altra strategia per la soluzione di questi problemi potrebbe essere quella di ricostruire la *library* utilizzando la strategia della *primer extension* che permetterebbe di ridurre la distanza della sequenza casuale a 7bp dalla box -35. Inoltre si potrebbe cercare di ottenere anche una maggior complessità per aumentare la probabilità di rappresentare la maggior parte dei motivi. Oppure si potrebbe valutare l'ipotesi della

costruzione di una *library* come quella utilizzata per gli omoeodomini (descritta nell'introduzione) che serve in casi in cui il TF ha una forza di legame al DNA troppo debole per cui non è in grado di attivare l'espressione dei geni *reporter*. Per risolvere il problema il TF viene prodotto come fusione sia alla subunità della RNA polimerasi sia ad un DBD parziale che serve per ancorare la proteina alla *library* poiché a monte della regione casuale è posto il motivo riconosciuto da questo DBD. Solo nel caso di interazione del fattore di trascrizione in esame si ottiene l'espressione dei geni *reporter*.

6. Analisi *in silico* dei fattori di trascrizione codificati dal genoma di *Laccaria bicolor*

La disponibilità della sequenza genomica del fungo basidiomicete simbiotico *L. bicolor* ha consentito di poter catalogare il repertorio dei fattori di trascrizione codificati dal genoma (tabella 4 in appendice). La classificazione *in silico* dei fattori di trascrizione è stata eseguita attraverso la catalogazione delle proteine in base alla predizione di domini di legame al DNA.

Per la classificazione ci si è basati sulle informazioni già presenti sul database di catalogazione dei fattori di trascrizione: *"DBD: Transcription factor prediction database"* (www.transcriptionfactor.org).

I dati preliminari ottenuti da questo database sono stati ulteriormente analizzati attraverso programmi bioinformatici per la predizione di domini conservati all'interno delle sequenze in esame. Inoltre sono stati ricercati i putativi fattori di trascrizione presenti nel genoma di *L. bicolor* omologhi a TF già caratterizzati in *S. cerevisiae* per cercare di attribuire una funzione ai TF isolati.

In *Laccaria,* come già evidenziato anche per *Tuber*, si riscontrano le classi maggiormente rappresentate nei funghi che sono bZIP, bHLH e gli *zinc finger.*

Questa analisi ci ha permesso di identificare un totale di 215 fattori di trascrizione codificati dal genoma di questo fungo basidiomicete. I fattori di trascrizione più abbondanti in *L. bicolor* sono gli *zinc finger* C2H2 (56) e gli *zinc cluster* (64). Insieme ricoprono il 56% di tutti i fattori di trascrizione catalogati *in silico*.

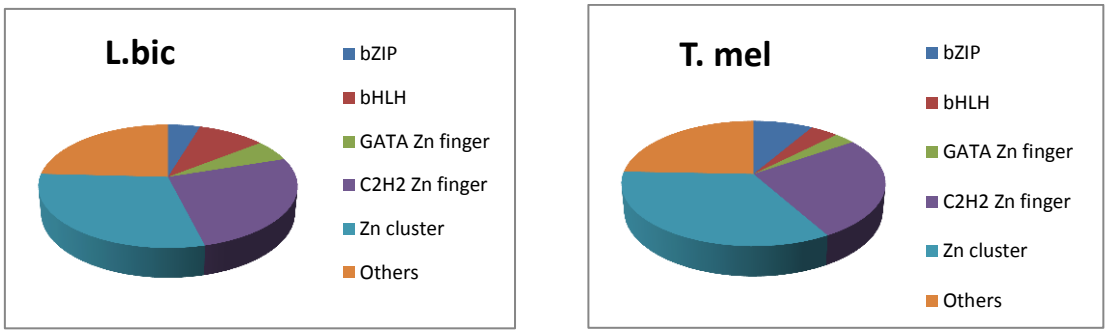


Fig.24 : Distribuzione dei fattori di trascrizione di *L. bicolor* e *T.melansoporum* all'interno delle famiglie maggiormente rappresentate nei funghi.

I genomi di *L. bicolor* e *T. melanosporum* codificano per un numero paragonabile di fattori di trascrizione, 215 il primo e 201 il secondo. Come si nota in figura la distribuzione dei C2H2 *zinc finger* e degli *zinc cluster* tra i due funghi simbiotici è molto simile mentre notiamo differenze tra i bHLH e i fattori GATA maggiormente abbondanti in *Laccaria* e i bZIP più rappresentanti in *Tuber*.

Inoltre in *L. bicolor* come accade anche in *Aspergillus* si trovano fattori di trascrizione che possiedono sia il dominio di legame C2H2 *zinc finger* che il dominio *zinc cluster*. Questa caratteristica risulta invece assente in *Tuber.*

Grazie all'analisi dell'omologia tra i fattori di trascrizione di *L. bicolor* e le proteine di lievito *S. cerevisiae* è stata possibile una predizione della funzione di alcuni TF presenti nel genoma del fungo. Questi TF possono essere suddivisi in gruppi in base alla funzione dei loro ipotetici geni bersaglio: sviluppo asessuale o sessuale, ciclo cellulare, metabolismo e risposta a stress o stimoli.

Per quanto riguarda i fattori di trascrizione coinvolti nel metabolismo vengono isolate proteine coinvolte nella biosintesi degli aminoacidi come Leu3 e Gcn4, proteine coinvolte nella repressione da catabolita, fattori di trascrizione con un ruolo nella gluconeogenesi e nel ciclo del gliossilato. Si isolano anche fattori di trascrizione coinvolti nell'assimilazione dello zolfo come Cbf1 e nel metabolismo del fosfato Pho2.

Emergono anche omologhi a proteine con funzione nello sviluppo. Ad esempio isoliamo il regolatore della morfologia cellulare Hms1 che promuove la crescita ifale e il regolatore dello sviluppo sessuale Mcm1 che è coinvolto nella regolazione dell'espressione genica in risposta a feromoni.

Sono presenti anche regolatori del ciclo cellulare e fattori di trascrizione coinvolti nella risposta a stress.

È in programma di integrare queste informazioni anche con i dati di espressione e di approfondire ulteriormente l'analisi di queste proteine per cercare di attribuire una funzione al maggior numero di TF estendendo le analisi di omologia anche ad altri funghi.

7. *Screening* genomico funzionale degli attivatori trascrizionali di *Laccaria bicolor e Populus trichocarpa*

Lo *screening* genomico funzionale condotto in *T. melanosporum* per l'identificazione degli attivatori trascrizionali è stato eseguito anche per le genoteche del fungo basidiomicete simbiotico *Laccaria bicolor* e del pioppo *Populus trichocarpa*, la pianta con cui *Laccaria* instaura la simbiosi.

Per identificare gli attivatori trascrizionali espressi nelle tre fasi del ciclo vitale di *L.bicolor*, sono state trasformate in lievito *S. cerevisiae* le genoteche FLM/FB di corpo fruttifero (FB) e micelio (FLM), ECM di ectomicorriza (ECM) e la genoteca ROOT della radice del pioppo.

La genoteca FLM/FB è stata prodotta con il cDNA di miceli cresciuti in condizioni diverse (carenza di azoto, carbonio, e fosforo) ed aggiunti al cDNA isolato da corpo fruttifero. Questo garantisce l'espressione di tutti i geni dato che si favorisce anche l'espressione di geni che vengono attivati in condizioni di carenza di nutrienti.

La genoteca FLM/FB possiede inserti di taglia media di 1,5 kb mentre la ECM e ROOT possiedono una taglia media di 1 kb.

Per la trasformazione del lievito sono stati utilizzati 20µg della genoteca FLM/FB e 15µg ciascuna delle genoteche ECM e ROOT. I trasformanti di ogni genoteca sono stati piastrati su 20 piastre contenenti 50 ml di terreno minimo privo di istidina e leucina, addizionato di 3-AT alla concentrazione 25mM in modo da poter selezionare trasformanti in grado di attivare la trascrizione del gene *reporter HIS3*.

Il numero di colonie piastrate totali è stato di $1,7*10^6$ per la genoteca FLM/FB, $2*10^6$ per ECM e $8*10^5$ per ROOT. Le colonie cresciute maggiormente sono state prelevate dalle piastre di trasformazione e fatte crescere in piastre SD-Leu in formato a 96 pozzetti per un giorno a 30°C. Sono state isolate 540 colonie per FLM/FB, 360 per ECM e 360 per ROOT. In ogni piastra sono stati aggiunti anche i controlli wt, m1, m2, pDEST32. Da queste piastre sono state assemblate piastre da 384 che sono state il punto di partenza per lo *screening* dell'attivazione dei geni *reporter*.

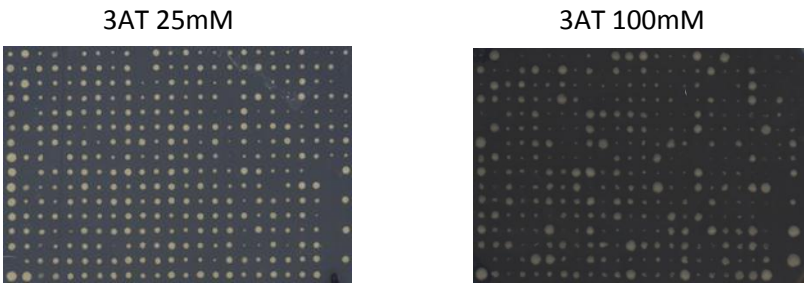


Fig. 25: Saggio di attivazione del gene *reporter HIS3* condotto sulle colonie appartenenti alla genoteca di micelio/corpo fruttifero di *L. bicolor*

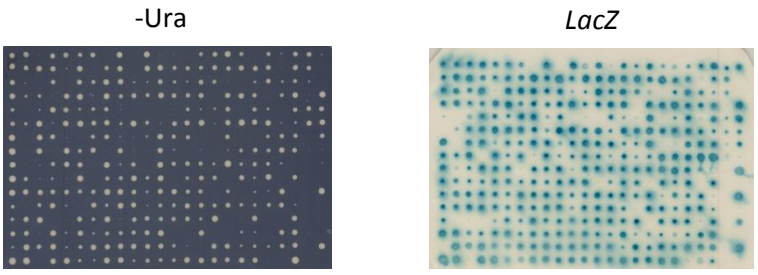


Fig. 26: Saggio di attivazione dei geni *reporter URA3* e *LacZ* condotto sulle colonie appartenenti alla genoteca di micelio/corpo fruttifero di *L. bicolor*

Analisi delle sequenze positive allo *screening* TAT

Come già descritto per *Tuber* alle colonie replicate sono stati assegnati dei valori di espressione dei geni *reporter.* Questi valori sono stati analizzati per identificare i cloni positivi ai due criteri stabiliti per isolare gli attivatori della trascrizione.

Le colonie risultanti positive ad entrambi i criteri sono: 222 colonie per FLM/FB (41% degli analizzati), 224 per ECM (62%) e 194 per ROOT (54%). Tutti questi cloni sono stati analizzati mediante PCR su colonia di lievito o mediante PCR su miniprep di lievito per ottenere gli ampliconi per il sequenziamento.

Si è ottenuta buona qualità di sequenziamento per un totale di 595 sequenze così suddivise:

- 195 dalla *library* di micelio/corpo fruttifero
- 213 dalla *library* di ectomicorriza;
- 187 dalla *library* di radice di pioppo.

Dopo le analisi bioinformatiche queste sequenze sono state organizzate in 83 *contigs* e 137 *singletons,* per un totale di 220 sequenze uniche di cui 80 di *Laccaria* e 140 di pianta. Le sequenze di pianta sono derivate dal contributo della genoteca di ectomicorriza e dalla genoteca di radice di pioppo.

	UNISEQU	CONTIG	SINGLETONS	TOTALE	FLM/FB	ECM	ROOT
WITH DBD	19	11	8	146	119	27	0
W/O DBD	61	20	41	113	76	37	0
TOT. FUNGO	80	31	49	259	195	64	0
PIANTA WITH DBD	61	37	24	223	0	99	124
PIANTA W/O DBD	79	15	64	113	0	50	63
TOT. PIANTA	140	52	88	336	0	149	187
TOT. PIANTA+FUNGO	220	83	137	595	195	213	187

Tabella 9 riassuntiva delle analisi delle sequenze di *Laccaria* e pioppo.

Delle 80 sequenze uniche di *Laccaria*, 19 sono rappresentate da fattori trascrizionali poiché contengono un *DNA binding domain* mentre le restanti 61 sequenze positive mancano di DBD noto. Tra queste si ritrovano 14 sequenze che vengono definite "*Nuclear protein and Unconventional activators*" (UAs) simili a geni che esprimono proteine con localizzazione nucleare o mista nucleare/citoplasmatica per le quali un ruolo diretto o indiretto nella regolazione della trascrizione è stato precedentemente dimostrato. Inoltre troviamo 4 sequenze che mostrano un dominio di legame ad acidi nucleici. Queste proteine sono buoni candidati per il ruolo di probabili fattori di trascrizione dato che vengono isolati mediante il TAT, hanno localizzazione nucleare e possiedono un dominio di legame ad acidi nucleici. Un'altra categoria di proteine nucleari sono quelle catalogate come secrete, nello *screening* TAT si isolano 5 proteine appartenenti a questa classe.

Tra le altre sequenze prive di un DBD 31 ricadono nella categoria "*Putative unconventional activators*" (PUAs) poiché in letteratura non è noto un loro ruolo nella trascrizione.

Le restanti 11 sequenze fanno parte delle proteine secrete prive di un segnale di localizzazione nucleare.

Dall'analisi delle 19 sequenze che posseggono un dominio di legame al DNA ritroviamo fattori di trascrizione coinvolti nella regolazione dello sviluppo sessuale del fungo come il fattore di trascrizione Pcc1 (Murata et al., 1998). Anche il fattore di trascrizione Prf1 è coinvolto nella regolazione dello sviluppo sessuale in particolare viene regolato da chinasi attivate da mitogeni e una volta attivo si lega a promotori dei

geni A e B del *mating type* del fungo e ne attiva la trascrizione (Brefort et al., 2005). Il fattore di trascrizione Ste12 è coinvolto nella regolazione dello sviluppo sessuale e nella formazione del corpo fruttifero ma anche della crescita ifale correlata alla penetrazione delle ife durante la colonizzazione dell'ospite (Wong Sak Hoi and Dumas, 2010).

Gcn4 è coinvolto nella regolazione della biosintesi degli amminoacidi. Dall'analisi ritroviamo anche il TF Chap1 che è coinvolto nella risposta a stress ossidativi (Lev et al., 2005). Inoltre vengono isolati anche fattori di trascrizione coinvolti nello sviluppo asessuale e nello sviluppo dei conidi come ad esempio Tec1 che è coinvolto nello sviluppo ifale.

È interessante notare che tre fattori di trascrizione ricadono in contig al cui interno ritroviamo un alto numero di sequenze: per Gcn4 troviamo un totale di 68 sequenze, per un fattore di trascrizione a funzione sconosciuta appartenente agli ZZ *zinc finger* isoliamo 21 sequenze e per un TF *ROX1-like* isoliamo 16 cloni. In tutti i casi la maggior parte delle sequenze isolate appartiene alla genoteca di micelio/corpo fruttifero.

Tra le sequenze che possiedono un DBD ritroviamo anche tre sequenze omologhe a repressori della trascrizione tra cui Rox1 e Rfx1, in questo caso possono essere isolati mediante lo *screening* per l'isolamento degli attivatori trascrizionali poiché all'interno del plasmide pDEST32 non viene clonata l'intera proteina ma soltanto un frammento della proteina che risulta in grado di attivare la trascrizione.

Esaminando i valori di espressione dei fattori di trascrizione isolati mediante TAT non sono stati isolati TF che mostrano una sovraespressione nel corpo fruttifero o nella ectomicorriza ad eccezione di due fattori di trascrizione coinvolti nello sviluppo asessuale (in grassetto nella tabella). In particolare l'omologo a Tec1 risulta essere maggiormente espresso di 25 volte in ectomicorriza e di 6 volte in corpo fruttifero rispetto al micelio mentre l'altro TF è 6 volte maggiormente espresso in ectomicorriza.

ANALISI TAT					BLASTP			
Seq. Name	**Protein ID**	**tot**	**FLM**	**ECM**	**prot ID**	**BlastP**	**Scer**	**DBD**
Contig05	293242	68	58	10	BAC55240.1	C-Gcn4	Gcn4	bZIP
ECM-L2_G03	**298274**	**1**	**0**	**1**	**XP_001388805.1**	**regulatory protein abaA**	**Tec1**	**TEA/ATTS**
ECM-L1_F11	307744	1	0	1	XP_002173151.1	meiotically up-regulated gene 152 protein		Myb
Contig62	386478	7	7	0	XP_001830477.1	PCC1		HMG1/HMG2
Contig41	393192	3	3	0	XP_001886200.1	STE12-like	Ste12	C2H2
ECM-L1_A03	457991	1	0	1	XP_002910064.1	NWD2		Forkhead
Contig40	458057	16	15	1	XP_001383328.2	ROX1-like HMG-box transcription factor		HMG1/HMG2
Contig65	481652	4	3	1	XP_003501812.1	RFX2	Rfx1	RFX
FLM-L2_D07	482609	1	1	0	XP_001828950.2	specific transcriptional repressor	Rox1	HMG1/HMG2
FLM-L2_D05	486090	1	1	0	XP_003034605.1	expressed protein		Zn2Cys6
ECM-L2_E05	626440	1	0	1	NP_587684.1	hypothetical protein		RNA polymerase I-specific transcription initiation factor RRN6-like
Contig46	628355	2	1	1	XP_567555.1	transcriptional regulatory protein		Transcription factor, fungi
FLM-L2_A08	633206	1	1	0	XP_001368548.1	*zinc finger* protein 850-like		C2H2
Contig43	640940	4	3	1	AAC32736.1	Prf1		HMG1/HMG2
Contig38	648888	9	6	3	AAC32736.1	Prf1		HMG1/HMG2
ECM-L1_H01	**656449**	**1**	**0**	**1**	**XP_001819986.2**	**regulatory protein abaA**		**TEA/ATTS**
Contig75	665554	2	2	0	AAS64313.1	Chap1	Yap1	bZIP
Contig07	682475	2	2	0	XP_001399919.1	C6 transcription factor (Mut3)		Zn2Cys6
Contig42	700637	21	16	5	EGO20236.1	hypothetical protein		ZZ *Zinc finger*

Tabella 10: Fattori di trascrizione di *Laccaria bicolor* isolati mediante TAT (anche tabella 5 in appendice)

Delle 61 proteine prive di un DBD (tabella 6 in appendice) ritroviamo 14 UAs tra cui Ede1, una proteina endocitotica, e l'ubiquitina ligasi che erano già state isolate dal TAT condotto in *T.melanosporum.* Vengono isolate anche la proteina Rad21 che è coinvolta nella coesione dei cromatidi durante la mitosi, Rad57 coinvolta nel meccanismo di riparazione del DNA e anche proteine nucleari coinvolte nella biogenesi dei ribosomi. Le proteine maggiormente interessanti e maggiormente promettenti per essere catalogate come fattori di trascrizione sono però 4 proteine di questa categoria che possiedono anche un dominio che si trova anche in proteine in grado di legare gli acidi nucleici. Non è possibile però introdurre tali sequenze nella categoria dei fattori di trascrizione poiché questi domini si riscontrano anche in proteine che non hanno un ruolo nella trascrizione.

Tra le 31 PUAs 19 sono a funzione sconosciuta e tra queste ritroviamo anche proteine uniche di *L. bicolor*. Tra le proteine a funzione nota ritroviamo trasportatori, proteine del perossisoma come la Pex19 isolata anche tra le proteine positive al TAT di *Tuber,*

proteine mitocondriali e proteine coinvolte nel trasporto mediato da vescicole e proteine chinasi. All'interno di questa categoria, come precedentemente descritto per *Tuber*, è possibile riscontrare sia dei falsi positivi a causa della metodologia utilizzata oppure proteine con una reale funzione nella trascrizione il cui ruolo non era ancora stato dimostrato.

Di particolare interesse sono le proteine secrete poiché candidate ad un ruolo di primo piano nella comunicazione con la pianta ospite durante l'instaurazione della simbiosi.

Con l'analisi TAT vengono isolate 16 proteine predette secrete di cui 5 a localizzazione nucleare che potrebbero essere coinvolte nella regolazione dell'espressione genica.

Tra le proteine secrete isolate quella espressa dalla sequenza che ricade nel contig64 ha un segnale di localizzazione extra-citoplasmatico all'N-terminale ed un segnale di localizzazione nucleare all'interno della sequenza. L'NLS ricade all'interno di una regione di dimerizzazione coiled-coil, inoltre si ritrovano anche residui di due leucine coinvolti nell'endocitosi mediata da vescicole di clatrina. Si ipotizza che la proteina venga secreta dalla cellula e che subisca un taglio proteolitico del peptide di secrezione che ne permette la dimerizzazione. La variazione conformazionale subita potrebbe causare l'esposizione del segnale di localizzazione nucleare e favorirne la la traslocazione finale all'interno del nucleo. La proteina dopo essere secreta viene nuovamente internalizzata, non è noto se in cellule vegatali o fungine.

Dato il grande interesse che suscitano le proteine secrete che risultano avere anche un ruolo come attivatori della trascrizione sono state saggiate due di queste proteine mediante *nuclear transportation trap*. Nessuna delle due proteine attiva l'spressione dei geni *reporter HIS3* e *LacZ*. Il saggio potrebbe dare risultato negativo poiché nel contesto eterologo mancano interattori che modificando o interagendo con la proteina ne causano la traslocazione nel nucleo.

Analizzando i dati di espressione delle proteine prive di DBD non emergono proteine differenzialmente espresse in uno stadio del ciclo vitale.

Per quanto riguarda invece i risultati del TAT per le sequenze di pianta ottenute dalla genoteca della radice di pioppo e da quella di ectomicorriza otteniamo 140 sequenze uniche suddivise in 61 proteine contenenti un DBD (tabella 7 in appendice) e 79 proteine prive di un DBD noto (26 UAs e 53 PUAs) (tabella 8 in appendice).

Tra le proteine che possiedono un dominio di legame aI DNA ritroviamo i fattori di trascrizione CONSTANS che sono proteine coinvolte nella regolazione dei processi di fioritura, della percezione della luce e della regolazione del ritomo circadiano (Wenkel

et al., 2006) e i Cbf che sono TF che possiedono un ruolo nel processo di acclimatazione per la resistenza al freddo (Welling and Palva, 2008).
Dall'analisi emergono anche fattori di trascrizione regolati dagli ormoni, ad esempio i fattori di trascrizione che vengono attivati dalle auxine.
I fattori di trascrizione contenenti un dominio NAC possono avere un ruolo nella risposta a stress biotici come l'infezione da patogeno, nella risposta a stress abiotici come shock termici e disidratazione (Olsen et al., 2005). I fattori di trascrizione ERF (*ethylene-responsive-factor*) sono responsabili dell'espressione genica in risposta a freddo, siccità, patogeni o a trattamenti con etilene, acido salicilico o altre molecole (Singh, 2002).
I fattori di trascrizione con un dominio MYB rappresentano una delle maggiori famiglie di TF in pianta e sono coinvolti in numerosi processi come la morfogenesi e lo sviluppo, la percezione della luce e la risposta a stress (Stracke, 2001).
Alcuni membri delle famiglie ERF, bZIP, WRKY e MYB sono gli attivatori trascrizionali maggiormente caratterizzati come responsabili delle risposte di difesa della pianta (Hegedus et al., 2003).
Anche nel caso di *Laccaria* non si tratta di una interazione tra patogeno e pianta ma di una interazione simbiotica. Questi fattori di trascrizione potrebbero quindi intervenire nella regolazione di processi collegati all'instaurazione della simbiosi e della comunicazione tra i due organismi.
I fattori di trascrizione appartenenti alla classe GRAS sono coinvolti nello sviluppo delle radici e nell'instaurazione della simbiosi. Ad esempio alcuni fattori di trascrizione sono coinvolti nella riprogrammazione dell'espressione genica in risposta a stimoli derivanti dall'organismo con cui instaurano la simbiosi (Hirsch and Oldroyd, 2009).
Tra i fattori di trascrizione che possiedono una funzione nota viene isolata anche la proteina LIGULELESS1 che è coinvolta nell'organogenesi fogliare (Moreno et al., 1997).
Tra le proteine nucleari e gli attivatori non convenzionali ritroviamo, come nel caso di *Laccaria*, 4 sequenze (in grassetto nella tabella) che possiedono un dominio di legame agli acidi nucleici. Questi domini si ritrovano spesso in proteine in grado di legare l'RNA. Queste proteine rientrano comunque in quelle sequenze candidate ad avere un ruolo nella trascrizione dato che vengono isolate dal saggio, sono nucleari e possiedono un dominio di legame ad acidi nucleici.
Tra le UAs ritroviamo subunità del complesso 26S del proteasoma per cui è nota una localizzazione mista nucleare e citoplasmatica. La loro localizzazione nucleare dipende

dal tipo cellulare e dalle condizioni di crescita. Si è dimostrata la localizzazione nucleare sia per la subunità 20S che per la subunità regolativa 19S di cui è formato il complesso 26S del proteasoma, così come posseggono localizzazione nucleare anche le proteine coinvolte nella reazione di ubiquitinazione delle proteine (von Mikecz, 2006). Inoltre è emerso da studi di letteratura che le proteine che fanno parte del complesso del proteasoma hanno caratteristiche di proteine *moonlighting* che posseggono un ruolo nel processo di trascrizione della RNA polimerasi III (Jeffery, 2003). È importante sottolineare che dall'analisi TAT di pioppo ma anche in quella di *Tuber* e *Laccaria* emerge la proteina E3 ubiquitina-ligasi come potenziale attivatore della trascrizione.

Anche il fattore di allungamento EF1B che svolge un ruolo nel processo di traduzione è stato identificato come proteina *moolighting*. Questa proteina svolge numerosi ruoli all'interno della cellula tra cui trasduzione del segnale, organizzazione del citoscheletro, apoptosi e possiede anche funzioni nucleari coinvolte nella sintesi dell'RNA o nella mitosi. Per questa proteina è stata dimostrata anche la sua capacità di interagire con un fattore di trascrizione la cui funzione rimane per ora sconosciuta (Ejiri, 2002).

Tra le proteine coinvolte in modo indiretto nella regolazione genica è presente anche SIZ1, coinvolta nella sumoilazione, per cui è stata dimostrata sia la sua localizzazione nucleare ma anche un suo ruolo nella regolazione di fattori di trascrizione con un ruolo nella risposta a stress come la siccità o il freddo oppure nella risposta a stimoli ormonali (Catala et al., 2007).

Nuclear proteins and unconventional activators (UAs)					
Seq. Name	**tot**	**ECM**	**ROOT**	**prot ID**	**description**
Contig02	5	1	4	XP_002533858.1	26S proteasome non-atpase regulatory subunit
Contig17	2	1	1	NP_196773.1	Ubiquitin-associated/translation elongation factor EF1B
Contig21	3	3	0	NP_193405.1	60S ribosomal protein L15-1
Contig33	2	1	1	AAF34804.1	CDK-activating kinase
Contig50	2	0	2	NP_180260.1	Ubiquitin-associated/translation elongation factor EF1B
Contig52	2	0	2	XP_002280700.1	protein LTV1 homolog
Contig59	9	7	2	XP_002275593.1	importin subunit alpha-1
Contig60	2	0	2	XP_003595566.1	CTD small phosphatase-like protein
Contig83	2	1	1	XP_002528348.1	o-linked n-acetylglucosamine transferase
ECM-L1_B08	1	1	0	XP_002888582.1	proteasome maturation factor UMP1 family
ECM-L1_G09	1	1	0	XP_002514774.1	cohesin subunit rad21
ECM-L1_H10	1	1	0	XP_002283831.2	26S protease regulatory subunit 6A homolog A
ECM-L2_B04	1	1	0	XP_002271062.1	E3 ubiquitin-protein ligase RING1-like
ECM-L2_D07	1	1	0	XP_002532142.1	transcription elongation factor s-II
ROOT-L1_E11	1	0	1	XP_002532142.1	transcription elongation factor s-II
ECM-L2_H11	1	1	0	XP_002510837.1	Transcriptional corepressor SEUSS
ROOT-L1_C04	1	0	1	NP_565199.1	DNA topoisomerase 2-associated protein PAT1
ROOT-L1_F06	1	0	1	XP_002284945.1	E3 SUMO-protein ligase SIZ1
ROOT-L1_F10	1	0	1	BAF46306.1	ataxin-2 related protein
ROOT-L1_G07	1	0	1	XP_002521404.1	DNA repair helicase rad5,16,
ROOT-L2_G05	1	0	1	XP_003630417.1	Serine/threonine protein kinase
ROOT-L1_D06	1	0	1	XP_002523135.1	kinesin heavy chain
ECM-L2_F03	1	1	0	XP_002511999.1	**RNA binding protein**
ROOT-L2_E05	1	0	1	XP_002279792.2	**protein MEI2-like 2-like**
ECM-L1_E02	1	1	0	XP_002279202.2	***zinc finger* CCCH domain-containing protein**
ROOT-L2_D04	1	0	1	XP_002268751.1	**pumilio homolog 5**

Tabella 11 : UAs isolate dallo *screening* TAT sulle genoteche di pioppo ed ectomicorriza. In grassetto sono evidenziate le proteine con un dominio di legame al DNA.

Tra i *putative unconventional activators* (tabella 8 in appendice) isoliamo Rin13 che è coinvolta nel processo di resistenza ad infezioni batteriche (Al-daoude et al., 2005) così come la proteina Sgt1, entrambe sono in grado di riconoscere proteine effettrici dei patogeni e di attivare la risposta immunitaria della pianta (Azevedo et al., 2006).

Emergono dalle analisi TAT anche proteine con ruolo nel metabolismo come le proteine che legano l'oxisterolo che hanno un ruolo nel trasporto intracellulare degli steroli oppure nel metabolismo di queste molecole.

Vengono isolate anche proteine come chinasi e fosfatasi che sono coinvolte nel meccanismo di trasduzione del segnale indotti da ormoni vegetali come etilene e acido abscissico.

Nell'analisi isoliamo anche la proteina di membrana con attività chinasica THESEUS1 che funge da recettore di stimoli esterni che causano modificazioni della parete cellulare vegetale. Questa proteina può essere attivata anche a seguito di un attacco da parte di un patogeno e la sua attivazione porta a sua volta alla regolazione dell'espressione di geni coinvolti nel meccanismo di difesa della pianta (Hématy et al., 2007).

Come per le proteine isolate in *Tuber* ritroviamo proteine coinvolte nei processi di trasporto o di autofagia. Isoliamo anche l'epossido idrolasi un enzima che nelle piante oltre ad avere una funzione nella biosintesi della cutina sembra ricoprire un ruolo anche nella produzione di metaboliti coinvolti nella difesa dagli attacchi di fungi patogeni (Morisseau and Hammock, 2005).

Infine la chinasi citoplasmatica SRK2A è coinvolta nella meccanismo di trasduzione del segnale in condizioni di stress ambientali come siccità o alta osmolarità (Thapa et al., 2011).

.

Conclusioni

In questo studio si è riusciti a mettere a punto una buona strategia per l'isolamento degli attivatori della trascrizione e per la catalogazione dei fattori di trascrizione grazie al recente sequenziamento di due funghi filamentosi simbiotici: *Tuber melanosporum* e *Laccaria bicolor.*

Le tecniche a cui ci si è affidati per la selezione e la successiva identificazione degli attivatori trascrizionali sono quelle basate sull'analisi del livello di espressione dei tre geni *reporter HIS3*, *URA3* e *LacZ* mediante due metodi che permettono di isolare solo le sequenze maggiormente promettenti. L'analisi di omologia ha permesso di identificare un buon numero di fattori di trascrizione tra cui alcuni tipici dei funghi, altri coinvolti nella regolazione di geni coinvolti nella resistenza a stress ossidativi o ai metalli pesanti e altri coinvolti nel metabolismo di azoto e carbonio. Inoltre sono stati isolati anche fattori di trascrizione codificati dal genoma della pianta che sono coinvolti nell'interazione pianta-fungo. Il TAT inoltre ha consentito la validazione funzionale di un quinto dei fattori di trascrizione catalogati *in silico* per quanto riguarda *Tuber melanosporum.*

Inoltre anche tra le proteine catalogate come ipotetiche si sono evidenziati dei domini attribuibili ad attivatori trascrizionali o a proteine a localizzazione nucleare. In particolare nello studio condotto in *Laccaria* emergono proteine a funzione sconosciuta che possiedono domini di legame ad acidi nucleici. Tra le proteine a localizzazione nucleare identificate dal nostro studio ne emergono alcune individuate anche in studi di proteomica eseguiti su organismi diversi ma sempre mirati all'isolamento delle proteine nucleari.

L'isolamento di alcune proteine che comunemente non hanno localizzazione nucleare ha indotto a pensare che la loro espressione nucleare fosse forzata dal trasporto della proteina nel nucleo. Per questo motivo si sono scelte alcune di queste sequenze per saggiarle mediante l'esperimento di *Nuclear Transportation Trap*.

Il risultato positivo ottenuto per la 3'-fosfoadenosina fosfosolfato (PAPS) reduttasi e le successive analisi che ci hanno permesso l'identificazione di un putativo dominio di attivazione mettono in evidenza come, grazie a questo *screening,* si possano isolare nuovi attivatori della trascrizione che potrebbero appartenere alla categoria delle *moonlighting protein* cioè a quelle proteine multifunzionali in grado di svolgere diverse funzioni anche in diversi comparti cellulari in modo indipendente.

È inoltre indispensabile sottolineare come tale metodo possa essere esteso all'analisi degli attivatori della trascrizione codificati dal genoma di qualsiasi organismo. Inoltre il TAT può essere utilizzato anche per eseguire esperimenti mirati se si è interessati ad attribuire o a verificare la funzione di attivatore di una particolare proteina.

Inoltre, le sequenze di particolare interesse che emergeranno da questi *screening* possono essere direttamente usate per l'identificazione dei loro interattori, dal momento che sia il vettore, sia il lievito in cui sono contenute sono adatti per effettuare uno *screening* di doppio ibrido.

Si è cercato di mettere a punto anche una metodologia che consentisse l'isolamento dei motivi di legame dei fattori di trascrizione codificati dal genoma di *Tuber* maggiormente interessanti perché preferenzialmente espressi nel corpo fruttifero o nell'ectomicorriza. A questo proposito si sono ottenuti risultati solamente per due delle proteine analizzate. Sarà necessario in futuro cercare di apportare le corrette modificazioni alla tecnica del *Bacterial one hybrid* in modo da riuscire ad ottenere informazioni dal maggior numero possibile di proteine prese in esame. Se questo non fosse possibile sarà necessario lo sviluppo di un'altra tecnica che consenta il raggiungimento dello scopo.

Inoltre sarà opportuna anche la messa a punto di una strategia che ci consenta una facile predizione e localizzazione dei motivi di legame di un fattore di trascrizione sul genoma in modo da poter sfruttare al meglio le informazioni ottenute dal B1H.

Bibliografia

Al-daoude, A., Zabala, M.D.T., Ko, J.-h. and Grant, M. (2005) RIN13 Is a Positive Regulator of the Plant Disease Resistance Protein RPM1. *Plant Disease*, 17, 1016-1028.

Amoutzias, G.D., Veron, a.S., Weiner, J., Robinson-Rechavi, M., Bornberg-Bauer, E., Oliver, S.G. and Robertson, D.L. (2007) One billion years of bZIP transcription factor evolution: conservation and change in dimerization and DNA-binding site specificity. *Molecular biology and evolution*, 24, 827-835.

Aro, N., Pakula, T. and Penttilä, M. (2005) Transcriptional regulation of plant cell wall degradation by filamentous fungi. *FEMS microbiology reviews*, 29, 719-739.

Azevedo, C., Betsuyaku, S., Peart, J., Takahashi, A., Noël, L., Sadanandom, A., Casais, C., Parker, J. and Shirasu, K. (2006) Role of SGT1 in resistance protein accumulation in plant immunity. *The EMBO journal*, 25, 2007-2016.

Ballario, P. and Macino, G. (1997) White collar proteins : PASsing the light signal in Neurospora crassa. *Trends in Microbiology*, 7123-7126.

Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Iii, E., P.w and Bulyk, M.L. (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol*, 24, 1429-1435.

Blumenstein, A., Vienken, K., Tasler, R., Purschwitz, J., Veith, D., Frankenberg-Dinkel, N. and Fischer, R. (2005) The Aspergillus nidulans phytochrome FphA represses sexual development in red light. *Current biology : CB*, 15, 1833-1838.

Borkovich, K.A., Alex, L.A., Yarden, O., Turner, G.E., Read, N.D., Seiler, S., Bell-pedersen, D., Paietta, J., Plesofsky, N., Plamann, M., Goodrich-tanrikulu, M., Schulte, U., Mannhaupt, G., Nargang, F.E., Radford, A., Selitrennikoff, C., Galagan, J.E., Dunlap, J.C., Loros, J.J., Catcheside, D., Inoue, H., Aramayo, R., Polymenis, M., Selker, E.U., Sachs, M.S., Marzluf, G.A., Paulsen, I., Ebbole, D.J., Zelter, A., Kalkman, E.R., Rourke, R.O., Bowring, F., Yeadon, J., Suzuki, K., Sakai, W., Pratt, R. and Freitag, M. (2004) Lessons from the Genome Sequence of Neurospora crassa : Tracing the Path from Genomic Blueprint to Multicellular Organism Lessons from the Genome Sequence of Neurospora crassa : Tracing the Path from Genomic Blueprint to Multicellular Organism. *Society*.

Brefort, T., Müller, P., Kahmann, R. and Mu, P. (2005) The High-Mobility-Group Domain Transcription Factor Rop1 Is a Direct Regulator of prf1 in Ustilago maydis The High-Mobility-Group Domain Transcription Factor Rop1 Is a Direct Regulator of prf1 in Ustilago maydis. *Society*, 4.

Bulyk, M.L. (2005) Discovering DNA regulatory elements with bacteria. *Nature biotechnology*, 23, 942-944.

Burley, S.K., Xie, X., Clark, K.L. and Shu, F. (1997) Histone-like transcription factors in eukaryotes. *Current opinion in structural biology*, 7, 94-102.

Busch, S., Eckert, S.E., Krappmann, S. and Braus, G.H. (2004) The COP9 signalosome is an essential regulator of development in the filamentous fungus Aspergillus nidulans. *Molecular Microbiology*, 49, 717-730.

Catala, R., Ouyang, J., Abreu, I.a., Hu, Y., Seo, H., Zhang, X. and Chua, N.-H. (2007) The Arabidopsis E3 SUMO ligase SIZ1 regulates plant growth and drought responses. *The Plant cell*, 19, 2952-2966.

Choi, J., Huh, S.U., Kojima, M., Sakakibara, H., Paek, K.-h. and Hwang, I. (2010) Article The Cytokinin-Activated Transcription Factor ARR2 Promotes Plant Immunity via TGA3 / NPR1-Dependent Salicylic Acid Signaling in Arabidopsis. *Developmental Cell*, 19, 284-295.

Cyert, M.S. (2001) Regulation of nuclear localization during signaling. *The Journal of biological chemistry*, 276, 20805-20808.

Deppmann, C.D., Alvania, R.S. and Taparowsky, E.J. (2006) Cross-species annotation of basic leucine zipper factor interactions: Insight into the evolution of closed interaction networks. *Molecular biology and evolution*, 23, 1480-1492.

Ejiri, S.-i. (2002) Moonlighting functions of polypeptide elongation factor 1: from actin bundling to zinc finger protein R1-associated nuclear localization. *Bioscience, biotechnology, and biochemistry*, 66, 1-21.

Etxebeste, O., Ni, M., Garzia, A., Kwon, N.-J., Fischer, R., Yu, J.-H., Espeso, E.a. and Ugalde, U. (2008) Basic-zipper-type transcription factor FlbB controls asexual development in Aspergillus nidulans. *Eukaryotic cell*, 7, 38-48.

Fassler, J., Landsman, D., Acharya, A., Moll, J.R., Bonovich, M. and Vinson, C. (2002) B-ZIP Proteins Encoded by the Drosophila Genome : Evaluation of Potential Dimerization Partners. *Genome Research*, 1190-1200.

Gajiwala, K.S. and Burley, S.K. (2000) Winged helix proteins. *Current opinion in structural biology*, 10, 110-116.

Gancedo, C. and Flores, C.-l. (2008) Moonlighting Proteins in Yeasts. *Society*, 72, 197-210.

Hall, D.A., Zhu, H., Zhu, X., Royce, T., Gerstein, M. and Snyder, M. (2009) Regulation of Gene Expression by a Metabolic Enzyme. *October*, 482.

Hanover, J.a., Love, D.C. and Prinz, W.a. (2009) Calmodulin-driven nuclear entry: trigger for sex determination and terminal differentiation. *The Journal of biological chemistry*, 284, 12593-12597.

Hardy, J.a. and Nelson, H.C. (2000) Proline in alpha-helical kink is required for folding kinetics but not for kinked structure, function, or stability of heat shock transcription factor. *Protein science : a publication of the Protein Society*, 9, 2128-2141.

Hegedus, D., Yu, M., Baldwin, D., Gruber, M., Sharpe, A., Parkin, I., Whitwill, S. and Lydiate, D. (2003) Molecular characterization of Brassica napus NAC domain transcriptional activators induced in response to biotic and abiotic stress. *Plant molecular biology*, 53, 383-397.

Hématy, K., Sado, P.-E., Van Tuinen, A., Rochange, S., Desnos, T., Balzergue, S., Pelletier, S., Renou, J.-P. and Höfte, H. (2007) A receptor-like kinase mediates the response of Arabidopsis cells to the inhibition of cellulose synthesis. *Current biology : CB*, 17, 922-931.

Hirsch, S. and Oldroyd, G.E.D. (2009) GRAS-domain transcription factors that regulate plant development. *Plant signaling & behavior*, 4, 698-700.

Hoffmann, B., Valerius, O., Andermann, M. and Braus, G.H. (2001) Transcriptional autoregulation and inhibition of mRNA translation of amino acid regulator gene cpcA of filamentous fungus Aspergillus nidulans. *Molecular biology of the cell*, 12, 2846-2857.

Hu, S., Xie, Z., Onishi, A., Yu, X., Jiang, L., Lin, J., Rho, H.-s., Woodard, C., Wang, H., Jeong, J.-s., Long, S., He, X. and Wade, H. (2009) Profiling the Human Protein-DNA Interactome Reveals ERK2 as a Transcriptional Repressor of Interferon Signaling. *Cell*, 139, 610-622.

Huberts, D.H.E.W. and van der Klei, I.J. (2010) Moonlighting proteins: an intriguing mode of multitasking. *Biochimica et biophysica acta*, 1803, 520-525.

Hudson, M.E. and Snyder, M. (2006) High-throughput methods of regulatory element discovery Tech Insight. *Biotechniques*, 41.

Jeffery, C.J. (1999) Moonlighting proteins. *Trends Biochem. Sci*, 24, 8-11.

Jeffery, C.J. (2003) Moonlighting proteins : old proteins learning new tricks. *Trends in genetics*, 19, 415-417.

Jeffery, C.J. (2009) Moonlighting proteins--an update. *Molecular bioSystems*, 5, 345-350.

Jones, S. (2004) An overview of the basic helix-loop-helix proteins. *Genome biology*, 5, 226.

Kernaghan, G. (2005) Mycorrhizal diversity: Cause and effect? *Pedobiologia*, 49, 511-520.

Kim, J.-W. and Dang, C.V. (2005) Multifaceted roles of glycolytic enzymes. *Trends in biochemical sciences*, 30, 142-150.

Kim, M.C., Chung, W.S., Yun, D.-j. and Cho, M.J. (2009) Calcium and Calmodulin-Mediated Regulation of Gene Expression in Plants. *Access*, 2.

Krappmann, S., Jung, N., Medic, B., Busch, S., Prade, R.a. and Braus, G.H. (2006) The Aspergillus nidulans F-box protein GrrA links SCF activity to meiosis. *Molecular microbiology*, 61, 76-88.

Lange, A., Mills, R.E., Lange, C.J., Stewart, M., Devine, S.E. and Corbett, A.H. (2007) Classical nuclear localization signals: definition, function, and interaction with importin alpha. *The Journal of biological chemistry*, 282, 5101-5105.

Lee, S.M., Kim, J.H., Cho, E.J. and Youn, H.D. (2009) A nucleocytoplasmic malate dehydrogenase regulates p53 transcriptional activity in response to metabolic stress. *Cell death and differentiation*, 16, 738-748.

Lev, S., Hadar, R., Amedeo, P., Baker, S.E., Yoder, C., Horwitz, B.A. and Yoder, O.C. (2005) Activation of an AP1-Like Transcription Factor of the Maize Pathogen Cochliobolus heterostrophus in Response to Oxidative Stress and Plant Signals Activation of an AP1-Like Transcription Factor of the Maize Pathogen Cochliobolus heterostrophus in Response. *Society*.

Liu, X., Noll, D.M., Lieb, J.D. and Clarke, N.D. (2005) DIP-chip: rapid and accurate determination of DNA-binding specificity. *Genome research*, 15, 421-427.

Luscombe, N.M., Austin, S.E., Berman, H.M. and Thornton, J.M. (2000) An overview of the structures of protein-DNA complexes. *Genome biology*, 1, REVIEWS001.

Ma, J. and Ptashne, M. (1987) A new class of yeast transcriptional activators. *Cell*, 51, 113-119.

Macara, I.A.N.G. (2001) Transport into and out of the Nucleus. *Society*, 65, 570-594.

MacPherson, S., Larochelle, M. and Turcotte, B. (2006) A fungal family of transcriptional regulators: the *zinc cluster* proteins. *Microbiology and molecular biology reviews : MMBR*, 70, 583-604.

Malik, S., Shukla, A., Sen, P. and Bhaumik, S.R. (2009) The 19 s proteasome subcomplex establishes a specific protein interaction network at the promoter for stimulated transcriptional initiation in vivo. *The Journal of biological chemistry*, 284, 35714-35724.

Markova, N.G., Pinkas-sarafova, A. and Simon, M. (2006) A Metabolic Enzyme of the Short-Chain Dehydrogenase / Reductase Superfamily May Moonlight in the Nucleus as a Repressor of Promoter Activity. *Journal of Investigative Dermatology*, 2019-2031.

Marshall, K.S., Zhang, Z., Curran, J., Derbyshire, S. and Mymryk, J.S. (2007) An improved genetic system for detection and analysis of protein nuclear import signals. *BMC Molecular Biology*, 3.

Martchenko, M., Levitin, A. and Whiteway, M. (2007) Transcriptional activation domains of the Candida albicans Gcn4p and Gal4p homologs. *Eukaryotic cell*, 6, 291-301.

Martin, F., Aerts, a., Ahrén, D., Brun, a., Danchin, E.G.J., Duchaussoy, F., Gibon, J., Kohler, a., Lindquist, E., Pereda, V., Salamov, a., Shapiro, H.J., Wuyts, J., Blaudez, D., Buée, M., Brokstein, P., Canbäck, B., Cohen, D., Courty, P.E., Coutinho, P.M., Delaruelle, C., Detter, J.C., Deveau, a., DiFazio, S., Duplessis, S., Fraissinet-Tachet, L., Lucic, E., Frey-Klett, P., Fourrey, C., Feussner, I., Gay, G., Grimwood, J., Hoegger, P.J., Jain, P., Kilaru, S., Labbé, J., Lin, Y.C., Legué, V., Le Tacon, F., Marmeisse, R., Melayah, D., Montanini, B., Muratet, M., Nehls, U., Niculita-Hirzel, H., Oudot-Le Secq, M.P., Peter, M., Quesneville, H., Rajashekar, B., Reich, M., Rouhier, N., Schmutz, J., Yin, T., Chalot, M., Henrissat, B., Kües, U., Lucas, S., Van de Peer, Y., Podila, G.K., Polle, a., Pukkila, P.J., Richardson, P.M., Rouzé, P., Sanders, I.R., Stajich, J.E., Tunlid, a., Tuskan, G. and Grigoriev, I.V. (2008) The genome of Laccaria bicolor provides insights into mycorrhizal symbiosis. *Nature*, 452, 88-92.

Martin, F., Kohler, A. and Duplessis, S. (2007) Living in harmony in the wood underground: ectomycorrhizal genomics. *Current opinion in plant biology*, 10, 204-210.

Martin, F., Kohler, A., Murat, C., Balestrini, R., Coutinho, P.M., Jaillon, O., Montanini, B., Morin, E., Noel, B., Percudani, R., Porcel, B., Rubini, A., Amicucci, A., Amselem, J., Anthouard, V., Arcioni, S., Artiguenave, F., Aury, J.-M., Ballario, P., Bolchi, A., Brenna, A., Brun, A., Buée, M., Cantarel, B., Chevalier, G., Couloux, A., Da Silva, C., Denoeud, F., Duplessis, S., Ghignone, S., Hilselberger, B., Iotti, M., Marçais, B., Mello, A., Miranda, M., Pacioni, G., Quesneville, H., Riccioni, C., Ruotolo, R., Splivallo, R., Stocchi, V., Tisserant, E., Viscomi, A.R., Zambonelli, A., Zampieri, E., Henrissat, B., Lebrun, M.-H., Paolocci, F., Bonfante, P., Ottonello, S. and Wincker, P. (2010) Périgord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature*, 464, 1033-1038.

Martin, F. and Selosse, M.-A. (2008) The Laccaria genome: a symbiont blueprint decoded. *The New phytologist*, 180, 296-310.

Maston, G.a., Evans, S.K. and Green, M.R. (2006) Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics*, 7, 29-59.

Meng, X., Brodsky, M.H. and Wolfe, S.A. (2005) A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nature Biotechnology*, 23, 988-994.

Meng, X. and Wolfe, S.a. (2006) Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system. *Nature protocols*, 1, 30-45.

Montecchio, L. (2006) Il ruolo delle simbiosi ectomicorriziche nel benessere delle piante forestali. 16, 321-326.

Moreno, M.a., Harper, L.C., Krueger, R.W., Dellaporta, S.L. and Freeling, M. (1997) Liguleless1 Encodes a Nuclear-Localized Protein Required for Induction of Ligules and Auricles During Maize Leaf Organogenesis. *Genes & Development*, 11, 616-628.

Morisseau, C. and Hammock, B.D. (2005) Epoxide hydrolases: mechanisms, inhibitor designs, and biological roles. *Annual review of pharmacology and toxicology*, 45, 311-333.

Murata, Y., Fujii, M., Zolan, M.E. and Kamada, T. (1998) Morphogenesis in Coprinus cinereus. *Seven*, 1761, 1753-1761.

Nadal, D. and Posas, F. (2010) Multilayered control of gene expression by stress-activated protein kinases. *EMBO Journal*, 29, 4-13.

Nowrousian, M., Frank, S., Koers, S., Strauch, P., Weitner, T., Ringelberg, C., Dunlap, J.C., Loros, J.J. and Kück, U. (2007) The novel ER membrane protein PRO41 is essential for sexual development in the filamentous fungus Sordaria macrospora. *Molecular microbiology*, 64, 923-937.

Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D. and Brodsky, M.H. (2008a) Analysis of Homeodomain Specificities Allows the Family-wide Prediction of Preferred Recognition Sites. *Molecular Medicine*, 1277-1289.

Noyes, M.B., Meng, X., Wakabayashi, A., Sinha, S., Brodsky, M.H. and Wolfe, S.A. (2008b) A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. *Access*, 36, 2547-2560.

Olsen, A.N., Ernst, H.a., Leggio, L.L. and Skriver, K. (2005) NAC transcription factors: structurally distinct, functionally diverse. *Trends in plant science*, 10, 79-87.

Pilecka, I., Banach-Orlowska, M. and Miaczynska, M. (2007) Nuclear functions of endocytic proteins. *European journal of cell biology*, 86, 533-547.

Plett, J.M. and Martin, F. (2011) Blurred boundaries : lifestyle lessons from ectomycorrhizal fungal genomes. *Trends in Genetics*, 27, 14-22.

Poupon, V., Polo, S., Vecchi, M., Martin, G., Dautry-Varsat, A., Cerf-Bensussan, N., Di Fiore, P.P. and Benmerah, A. (2002) Differential nucleocytoplasmic trafficking between the related endocytic proteins Eps15 and Eps15R. *The Journal of biological chemistry*, 277, 8941-8948.

Rabanales, C.U.D. (2008) Ctf1 , a transcriptional activator of cutinase and lipase genes in Fusarium oxysporum is dispensable for virulence. *Molecular Plant Pathology*, 9, 293-304.

Ren, T., Qu, F. and Morris, T.J. (2000) HRT gene function requires interaction between a NAC protein and viral capsid protein to confer resistance to turnip crinkle virus. *The Plant cell*, 12, 1917-1926.

Rhee, Y., Gurel, F., Gafni, Y., Dingwall, C. and Citovsky, V. (2000) A genetic system for detection of protein nuclear import and export. *Nature Biotechnology*, 433-437.

Scazzocchio, C. (2000) The fungal GATA factors. *Current opinion in microbiology*, 3, 126-131.

Scherer, M., Wei, H., Liese, R. and Fischer, R. (2002) Aspergillus nidulans Catalase-Peroxidase Gene (cpeA) Is Transcriptionally Induced during Sexual Development through the Transcription Factor StuA Aspergillus nidulans Catalase-Peroxidase Gene (cpeA) Is Transcriptionally Induced during Sexual Developme. *Society*.

Schweizer, a., Rupp, S., Taylor, B.N., Röllinghoff, M. and Schröppel, K. (2000) The TEA/ATTS transcription factor CaTec1p regulates hyphal development and virulence in Candida albicans. *Molecular microbiology*, 38, 435-445.

Scott, E.M. and Pillus, L. (2010) Homocitrate synthase connects amino acid metabolism to chromatin functions through Esa1 and DNA damage. *Genes & Development*, 1903-1913.

Shelest, E. (2008) Transcription factors in fungi. *FEMS microbiology letters*, 286, 145-151.

Shore, P. and Sharrocks, a.D. (1995) The MADS-box family of transcription factors. *European journal of biochemistry / FEBS*, 229, 1-13.

Singh, K. (2002) Transcription factors in plant defense and stress responses. *Current Opinion in Plant Biology*, 5, 430-436.

Sriram, G., Martinez, J.A., Mccabe, E.R., Liao, J.C. and Dipple, K.M. (2005) Single-gene disorders: what role could moonlighting enzymes play. *Am. J. Hum. Genet*, 76, 911-924.

Stegmaier, P., Kel, A.E. and Wingender, E. (2004) Systematic DNA-binding domain classification of transcription factors. *Genome informatics. International Conference on Genome Informatics*, 15, 276-286.

Stracke, R. (2001) The R2R3-MYB gene family in Arabidopsis thaliana Ralf Stracke , Martin Werber and Bernd Weisshaar *. *Current Opinion in Plant Biology*, 447-456.

Strauss, J., Muro-pastor, M.I. and Scazzocchio, C. (1998) The Regulator of Nitrate Assimilation in Ascomycetes Is a Dimer Which Binds a Nonrepeated , Asymmetrical Sequence The Regulator of Nitrate Assimilation in Ascomycetes Is a Dimer Which Binds a Nonrepeated , Asymmetrical Sequence. *Microbiology*, 18.

Thapa, G., Dey, M., Sahoo, L. and Panda, S.K. (2011) An insight into the drought stress induced alterations in plants. *Biologia Plantarum*, 55, 603-613.

Thorogate, R. and Török, K. (2004) Ca 2 + -dependent and -independent mechanisms of calmodulin nuclear translocation. *Journal of Cell Science*, 2.

Titz, B.r., Thomas, S., Rajagopala, S.V. and Chiba, T. (2006) Transcriptional activators in yeast. *Time*, 34.

Toledo-ortiz, G., Huq, E. and Quail, P.H. (2003) The Arabidopsis Basic / Helix-Loop-Helix Transcription Factor Family. *Society*, 15, 1749-1770.

Tsang, H.T.H., Connell, J.W., Brown, S.E., Thompson, A., Reid, E. and Sanderson, C.M. (2006) A systematic analysis of human CHMP protein interactions : Additional MIT domain-containing proteins bind to multiple components of the human ESCRT III complex. *Genomics*, 88, 333-346.

Tuskan, G.a., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, a., Schein, J., Sterck, L., Aerts, a., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, a., Brunner, a., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.-L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J.,

Degroeve, S., Déjardin, a., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, a., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjärvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, a., Kalluri, U., Larimer, F., Leebens-Mack, J., Leplé, J.-C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, a., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouzé, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, a., Sterky, F., Terry, a., Tsai, C.-J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y. and Rokhsar, D. (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science (New York, N.Y.)*, 313, 1596-1604.

Ueki, N., Oda, T., Kondo, M., Yano, K., Noguchi, T. and Muramatsu, M.-a. (1998) Selection system for genes encoding nuclear-targeted proteins. *Nature Biotechnology*, 16.

Ugalde, U. (2006) Autoregolatory signals in mycelial fungi.pdf.

Vecchi, M., Polo, S., Poupon, V., van de Loo, J.W., Benmerah, a. and Di Fiore, P.P. (2001) Nucleocytoplasmic shuttling of endocytic proteins. *The Journal of cell biology*, 153, 1511-1517.

Vinson, C., Myakishev, M., Acharya, A., Mir, A.A., Moll, J.R. and Bonovich, M. (2002) MINIREVIEW Classification of Human B-ZIP Proteins Based on Dimerization Properties. *Asha*, 22, 6321-6335.

von Mikecz, A. (2006) The nuclear ubiquitin-proteasome system. *Journal of cell science*, 119, 1977-1984.

Walhout, A.J.M. (2006) Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome research*, 16, 1445-1454.

Welling, A. and Palva, E.T. (2008) Involvement of CBF transcription factors in winter hardiness in birch. *Plant physiology*, 147, 1199-1211.

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A. and Coupland, G. (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. *The Plant cell*, 18, 2971-2984.

Wolfe, S.A., Nekludova, L. and Pabo, C.O. (2000) DNA recognition by Cys2His2 zinc finger proteine. *Annual Reviews*, 3, 183-212.

Wong Sak Hoi, J. and Dumas, B. (2010) Ste12 and Ste12-like proteins, fungal transcription factors regulating development and pathogenicity. *Eukaryotic cell*, 9, 480-485.

Yang, V.W. (1998) Issues and Opinions in Nutrition Eukaryotic Transcription Factors : Identification , Characterization. *ReVision*, 2045-2051.

Ye, R., Yao, Q.-h., Xu, Z.-h. and Xue, H.-w. (2004) Development of an efficient method for the isolation of factors involved in gene transcription during rice embryo development. 348-357.

Yu, J.-H. (2006) Heterotrimeric G protein signaling and RGSs in Aspergillus nidulans. *Journal of microbiology (Seoul, Korea)*, 44, 145-154.

Yu, Z., Lemongello, D., Segel, I.H. and Fisher, A.J. (2008) Crystal structure of Saccharomyces cerevisiae 3'-phosphoadenosine-5'-phosphosulfate reductase complexed with adenosine 3',5'-bisphosphate. *Biochemistry*, 47, 12777-12786.

Zheng, B., Han, M., Bernier, M. and Wen, J.-k. (2009) Nuclear actin and actin-binding proteins in the regulation of transcription and gene expression. *The FEBS journal*, 276, 2669-2685.

Zheng, L., Roeder, R.G. and Luo, Y. (2003) S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. *Cell*, 114, 255-266.

Zimmermann, I.M., Heim, M.a., Weisshaar, B. and Uhrig, J.F. (2004) Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant journal : for cell and molecular biology*, 40, 22-34.

Appendice

Tabella 1. Fattori di trascrizione a funzione nota codificati dal genoma di *T.melanosporum*

Functional classes and gene names[a]	Putative gene product function[b]	Preferential expression[c]		
		ECM	FB	FLM
Cell wall				
TmelXlnR	Activator of plant cell wall degrading enzymes	↑		
TmelAce1	Repressor of plant cell wall degrading enzymes		↓	
TmelCtf1	Transcriptional activator of cutinase genes			
TmelMrgA	Regulation of cell wall integrity			
TmelRlmA	Maintenance of cell wall integrity			
Development				
Sexual development and fruiting body formation				
TmelStuA	Regulator of fruiting body and asexual development			
TmelSteA	Regulator of fruiting body development		↓	
TmelMat 1-2-1	MAT1-2-1 mating-type protein			
TmelMcm1	Regulator of pheromone response			
TmelSte11	Regulator of pheromone response			
TmelAsl1	Ascospore Lethal 1			
TmelFf7	Female Fertility	↑		
TmelSub2	SUBmerged protoperithecia			
TmelNosA	Number Of Sexual spores, regulator of sexual development			
TmelNdt80	Meiosis-specific transcription factor			
Asexual development and basal hyphal growth				
TmelLah3	Long Aerial Hyphae, regulator of asexual development		↓	
TmelTah1	Tall Aerial Hyphae, regulator of asexual development			
TmelCol27	COLonial, regulator of hyphal growth			
TmelCol21	COLonial, regulator of hyphal growth			
TmelVad5	Vegetative Asexual Development, regulator of hyphal growth and asexual development			
TmelVad3-1	Vegetative Asexual Development, regulator of hyphal growth and asexual development			↑
TmelVad3-2	Vegetative Asexual Development, regulator of hyphal growth and asexual development	↑		
TmelKal1	KALeidoscope-1, regulator of hyphal growth and asexual development			
TmelAabA	Regulator of conidiation			
TmelDevR	Required for conidiophore development	↑		
TmelFlbB	Fluffy Low BrlA expression, regulator of asexual and conidiophore development			
TmelFlbC	Fluffy Low BrlA expression, regulator of asexual and conidiophore development			
TmelFlbD	Fluffy Low BrlA expression, regulator of asexual and conidiophore	↑		

	development			
TmelOefC	Over Expressed Fluffy	↑		
TmelHms1	Regulator of cell morphology			
TmelCon7	Cell morphology regulator			
Other				
TmelAda1	All Development Altered, regulator of basal hyphal growth & asexual and sexual development			
TmelAda3	All Development Altered, regulator of basal hyphal growth & asexual and sexual development	↑		
Cell Cycle				
TmelSwi6	MBF complex, regulator of cell cycle			
TmelMbp1	MBF complex, regulator of cell cycle		↓	
TmelFkh2	Regulator of cell cycle			
TmelSep1	Regulator of cell cycle			
TmelSwi5	Regulator of cell cycle			
TmelYox1	Regulator of cell cycle			
TmelUme6	Key transcriptional regulator of early meiotic genes			
TmelRme1	Regulator of cell cycle: prevents meiosis, promotes mitosis	↑		
Metabolism				
Carbon				
TmelCreA	Major Carbon catabolite REpression protein			
TmelNrg1	Carbon Catabolite Repression			
TmelAzf1	Glucose responsive transcription factor		↓	
TmelRgt1	Glucose responsive transcription factor			
TmelFacB	Activates transcription of genes required for acetate utilization			
TmelAcuM	Positive regulator of gluconeogenesis			
TmelAcuK	Positive regulator of gluconeogenesis		↑	
Nitrogen				
TmelAreA	Major, positively acting, nitrogen regulatory protein		↓	
TmelTamA	Positively acting nitrogen regulatory protein		↓	
TmelAreB	Major nitrogen catabolite repression protein		↓	
TmelMeaB	Nitrogen catabolite repression			
TmelGlcD	Regulation of nitrogen metabolism			
TmelCpcA	Cross Pathway Control protein, amino acid metabolism; regulator of fruiting body development			
TmelNirA-1	Pathway specific, positively acting nitrate regulatory protein			
TmelNirA-2	Pathway specific, positively acting nitrate regulatory protein			
TmelNirA-3	Pathway specific, positively acting nitrate regulatory protein	↑		
TmelAro80	Pathway specific, positively acting aromatic amino acid regulatory protein			
TmelUay1	Pathway specific, positively acting purine regulatory protein		↑	
TmelAmdX	Pathway specific, positively acting acetamidase regulatory protein			

TmelArcA	Pathway specific, positively acting arginine regulatory protein	↑		
TmelPpr1	Pathway specific, positively acting pyrimidine regulatory protein			
TmelLys14	Activator of lysine biosynthesis pathway			
Sulfur				
TmelMet32	Activator of sulfur metabolism		↓	
TmelMetR	Activator of sulfur metabolism			
TmelCbf1	Activator of sulfur metabolism; Centromere Binding protein		↑	
Lipids				
TmelOaf1	Activator of fatty acid metabolism			
TmelFarA	Activator of short- and long-chain fatty acid metabolism			↓
TmelFarB	Activator of short-chain fatty acid metabolism		↑	
TmelUpc2	Sterol regulatory element binding protein			
Other				
TmelPalcA	Transcription factor, involved in regulation of phosphate acquisition		↓	
TmelHapB	CCAAT binding complex, subunit B	↑		
TmelHapC	CCAAT binding complex, subunit C			
TmelHapE	CCAAT binding complex, subunit E			
TmelHapX	CCAAT binding complex, subunit X; iron responsive factor			
Stress and stimuli response				
TmelSfl1	Activator of stress responsive genes			
TmelFcr3	FluConazole Resistance protein 3		↓	
TmelSfp1	Cell cycle regulator in response to nutrients and stress			
TmelAsg1	Regulator of stress response & drug resistance			
TmelRpn4	Regulator of response to stress			
TmelHsf1	Heat Shock transcription Factor			
TmelAp1	Regulator of oxidative stress tolerance			
TmelSkn7	Response to osmotic & oxidative stress			
TmelPacC	Activator of alkaline-induced genes; repressor of acid-induced genes			
TmelHaa1	Weak acid stress response regulator			
TmelCrz1	Activator of genes involved in stress response			
TmelZap1	Activator of zinc responsive genes	↑		
TmelHacA	Unfolded protein response			
TmelWc1	Light response and circadian rhythm regulator			
TmelWc2	Light response and circadian rhythm regulator			
TmalSltA	Response to abiotic stress regulator			↑

[a]fattori trascrizionali di *T. melanosporum* contenenti *DNA binding domain*, raggruppati nelle maggiori classi funzionali basati sull'omologia con TFs funzionalmente caratterizzati in altri funghi; il nome dei geni è basato dal nome dei corrispondenti omologhi.

[b]Funzioni putative specifiche dei TFs di *Tuber*, dedotte dalle funzioni dei loro omologhi caratterizzati.

[c]Le frecce indicano il rapporto di espressione, derivante dai dati oligoarray e RNA-seq, ≥ 2 (↑) o ≤ 0.5 (↓) in uno stadio del ciclo vitale rispetto agli altri due. (ECM, ectomicorriza; FB: corpo fruttifero; FLM: micelio); i TFs TAT-positivi sono evidenziati in grassetto.

Tabella 2: classificazione delle sequenze di pianta ottenute dallo *screening* TAT condotto sulla genoteca di ectomicorriza di *Tuber melanosporum*

seq ID	ECM	acc #	BLAST result	Organism	score	e-value
Transcription factors						
Contig02	2	ABQ96643.1	NAC domain protein	*Citrus sinensis*	140	2E-31
Contig04	4	XP_002321956.1	NAC domain protein. IPR003441	*Populus trichocarpa*	98,6	6E-19
Contig05	3	XP_002533146.1	DNA binding protein, putative	*Ricinus communis*	247	8E-64
Contig07	9	BAE48667.1	NAC family protein	*Prunus mume*	64,7	1E-08
Contig11	2	XP_002519841.1	transcription factor, putative	*Ricinus communis*	73,6	2E-11
Contig13	9	ACV92008.1	WRKY transcription factor 6	*Populus tomentosa x P. bolleana*	142	3E-32
Contig16	24	CAE54591.1	ethylene transcription factor	*Fagus sylvatica*	540	5E-152
E1BD-G1_A06	1	AAM65338.1	NAC, putative	*Arabidopsis thaliana*	121	7E-26
E1BD-G1_B06	1	XP_002512809.1	Ethylene-responsive transcription factor	*Ricinus communis*	127	5E-28
E1BD-G1_D01	1	XP_002284400.1	PREDICTED: similar to MYB transcription factor	*Vitis vinifera*	190	6E-59
E1BD-G1_D07	1	ABW34392.1	R2R3 Myb14 transcription factor	*Vitis vinifera*	282	7E-75
E1BD-G1_E02	1	BAC98494.1	AG-motif binding protein-4	*Nicotiana tabacum*	82,8	1E-14
E1BD-G1_E09	1	ABN08215.1	Zinc finger, CCCH-type	*Medicago truncatula*	277	4E-75
E1BD-G1_F11	1	NP_193346	ARR2 transcription factoR	*Arabidopsis thaliana*	87	6E-16
E1BD-G1_G01	1	XP_002525655.1	transcription factor, putative	*Ricinus communis*	347	2E-94
E1BD-G2_C05	1	ACV92023.1	WRKY transcription factor 21	*Populus tomentosa x P. bolleana*	293	5E-78
E1BD-G2_C12	1	XP_002511518.1	homeobox protein, putative	*Ricinus communis*	346	1E-103
E1BD-G2_D05	1	BAG55220.1	transcription factor	*Sesbania rostrata*	107	3E-22
E1BD-G2_D06	1	BAD01556.1	ERF-like protein	*Cucumis melo*	73,9	5E-12
E1BD-G2_E06	1	XP_002533810.1	nuclear receptor binding set domain containing protein 1	*Ricinus communis*	201	2E-50
E1BD-G2_G06	1	ACV92025.1	WRKY transcription factor 23	*Populus tomentosa x P. bolleana*	263	4E-69

E1BD-G2_H06	1	XP_002533753.1	transcription factor, putative	*Ricinus communis*	186	9E-46
Unconventional putative activators						
Contig03	6	XP_002509541.1	Polcalcin Jun o, putative	*Ricinus communis*	135	6E-30
Contig17	2	AF452452_1	serine acetyltransferase	*Glycine max*	68,6	7E-10
Contig18	2	XP_002513362.1	serine/threonine protein kinase, putative	*Ricinus communis*	121	7E-26
E1BD-G2_E10	1	XP_002301590.1	fasciclin-like arabinogalactan protein 12.2	*Populus trichocarpa*	269	6E-71
E1BD-G2_F08	1	NP_189233.1	transferase family protein	*Arabidopsis thaliana*	51,6	0,00003
Contig59	2	NP_187405.1	Calmodulin, ARF-GAP domain 11	*Arabidopsis thaliana*	100	6E-29
Protein with unknown function						
Contig06	2	XP_002266912.1	PREDICTED: hypothetical protein	*Vitis vinifera*	233	2E-59
E1BD-G1_A02	1	XP_002328555.1	predicted protein	*Populus trichocarpa*	77	6E-13
E1BD-G1_B03	1	XP_002269465.1	PREDICTED: hypothetical protein	*Vitis vinifera*	239	6E-79
E1BD-G1_G03	1	XP_002298427.1	predicted protein	*Populus trichocarpa*	133	7E-30
E1BD-G1_G09	1	XP_002523666.1	conserved hypothetical protein	*Ricinus communis*	282	1E-74
E1BD-G2_A10	1	XP_002273611.1	PREDICTED: hypothetical protein	*Vitis vinifera*	159	9E-38
E1BD-G2_C03	1	XP_002272815.1	PREDICTED: hypothetical protein	*Vitis vinifera*	118	2E-25
E1BD-G2_D09	1	XP_002275378.1	PREDICTED: hypothetical protein	*Vitis vinifera*	287	2E-76
E1BD-G2_D12	1	XP_002262623.1	PREDICTED: hypothetical protein	*Vitis vinifera*	62,8	1E-08
E1BD-G2_F01	1	XP_002276336.1	PREDICTED: hypothetical protein	*Vitis vinifera*	90,5	3E-36
E1BD-G2_F09	1	XP_002284183.1	PREDICTED: hypothetical protein	*Vitis vinifera*	290	4E-77
E1BD-G2_G08	1	CAN59796.1	hypothetical protein	*Vitis vinifera*	116	5E-33
E1BD-G2_H07	1	XP_002278968.1	PREDICTED: hypothetical protein	*Vitis vinifera*	142	1E-32

Tabella 3: Valore di espressione dei fattori di trascrizione maggiormente espressi in corpo fruttifero e ectomicorriza di *Tuber melanosporum* selezionati per l'analisi TAT mirata

Sequence information			RNA-seq analysis					Array analysis						
Gene model #	**DBD**	**Gene name**	**#R ECM**	**#R FB**	**#R FLM**	**ratio ECM/FLM**	**ratio FB/FLM**	**ECM**	**FB**	**FLM**	**ratio ECM/FLM**	**P value ECM/FLM**	**ratio FB/FLM**	**P value FB/FLM**
Conserved hypothetical protein (CH-TF)														
GSTUMT00004451001	homeobox	CHP_GSTUMT-04451001	233	2	9	48,8	0,2	9813,923	163,802	291,073	33,7	0,00	0,6	0,26
GSTUMT00009188001	*zinc cluster*	CHP_GSTUMT-09188001	7	450	10	1,3	37,0	148,711	2854,305	198,91	0,7	0,80	14,3	0,05
GSTUMT00000561001	*zinc cluster*	CHP_GSTUMT-00561001	38	4	13	5,5	0,3	1901,109	351,844	257,985	7,37	0,00	1,36	0,32
GSTUMT00004341001	*zinc cluster*	CHP_GSTUMT-04341001	256	232	61	7,9	3,1	4376,213	3597,443	970,152	4,51	0,02	3,71	0,09
GSTUMT00006234001	CP2	CHP_GSTUMT-06234001	234	412	88	5,0	3,9	5006,584	3075,767	919,364	5,45	0,01	3,35	0,10
GSTUMT00006533001	*zinc cluster*	CHP_GSTUMT-06533001	29	29	3	18,2	8,0	775,508	898,707	30,346	25,56	0,01	29,62	0,00
DBD-containing hypothetical protein (HP-DBD)														
GSTUMT00000279001	C2H2	HP_GSTUMT-00279001	169	66	35	9,1	1,6	3466,549	1456,899	848,15	4,1	0,01	1,7	0,26
GSTUMT00005836001	*zinc cluster*	HP_GSTUMT-05836001	312	467	46	12,8	8,4	12544,54	4706,4	2277,766	5,5	0,01	2,1	0,34
GSTUMT00009814001	*zinc cluster*	HP_GSTUMT-09814001	1	219	1	1,9	180,3	1,255	4478,21	1	1,3	0,51	4478,2	0,00
GSTUMT00011475001	Myb-like	HP_GSTUMT-11475001	303	28	70	8,2	0,3	9952,028	1353,288	1353,856	7,4	0,01	1,0	1,00
GSTUMT00012451001	*zinc cluster*	TmelXLNR-2	2304	4	151	28,8	0,0	37909,92	379,432	5322,868	7,1	0,00	0,1	0,00
GSTUMT00004776001	forkhead	HP_GSTUMT-04776001	0	1	0			42,93	433,347	5,129	8,37	0,35	84,49	0,00
Putative transcription factor (pTF)														
GSTUMT00000111001	*zinc cluster*	TmelACUK	24	504	64	0,7	6,5	160,574	5216,429	447,522	0,4	0,09	11,7	0,00
GSTUMT00000814001	bHLH	TmelCBF1	165	1632	96	3,2	14,0	7294,759	19172,46	3701,654	2,0	0,02	5,2	0,00
GSTUMT00008124001	*zinc cluster*	TmelFF7	255	31	52	9,2	0,5	5291,609	1183,519	986,377	5,4	0,00	1,2	0,51

Tabella 4: Fattori di trascrizione codificati dal genoma di *L.bicolor*

protein ID v2	gene ID v2	DBD	Sc ortholog	other domains
709955	estExt_fgenesh3_kg.C_190066	APSES	Mbp1	IPR002110: Ankyrin repeat; IPR020683: Ankyrin repeat-containing domain
694007	fgenesh3_kg.LG_10_#_520_#_Locus_8004_Transcript_4	APSES	Swi6	IPR002110: Ankyrin repeat
442607	gm1.7051_g	APSES	Xbp1	
455773	estExt_Genemark1.C_LG_2_t20248	APSES		
441960	gm1.6404_g	APSES		
706529	estExt_fgenesh3_kg.C_LG_8_t20029	bHLH	Cbf1	
583411	e_gw1.3.7825.1	bHLH	Hms1	IPR017956: AT hook, DNA-binding motif
663232	estExt_fgenesh1_pg.C_LG_6_t10254	bHLH		
294581	Lacbi1.estExt_fgenesh2_pg.C_160074	bHLH		
625677	estExt_Genewise1Plus.C_LG_2_t10281	bHLH		
679667	fgenesh3_kg.LG_1_#_571_#_Locus_8282_Transcript_6/	bHLH		
655850	estExt_fgenesh1_kg.C_LG_1_t30003	bHLH		
700814	estExt_fgenesh3_kg.C_LG_1_t50386	bHLH		
657026	estExt_fgenesh1_kg.C_LG_3_t10355	bHLH		
550021	gw1.1.12675.1	bHLH		
328297	Lacbi1.fgenesh3_pg.C_scaffold_16000062	bHLH		
568346	e_gw1.1.2961.1	bHLH		
576537	e_gw1.1.13509.1	bHLH		
622992	estExt_Genewise1Plus.C_LG_1_t120115	bHLH		
574270	e_gw1.1.406.1	bHLH		
328297	Lacbi1.fgenesh3_pg.C_scaffold_16000062	bHLH		
328369	Lacbi1.fgenesh3_pg.C_scaffold_16000134	bHLH		
607731	e_gw1.10.369.1	bHLH		IPR008906: HAT dimerisation; IPR012337: Ribonuclease H-like
328369	Lacbi1.fgenesh3_pg.C_scaffold_16000134	bHLH		
438987	gm1.3431_g	bHLH		
687441	fgenesh3_kg.LG_3_#_2715_#_Locus_1269_Transcript_1	bHLH		
653451	estExt_Genewise1.C_12_t30363	bZIP	Cin5	
293242	Lacbi1.estExt_fgenesh2_pg.C_50565	bZIP	Gcn4	

574778	e_gw1.1.1973.1	bZIP	Yap3	IPR018287: Hap4 transcription factor, heteromerisation domain
445028	gm1.9472_g	bZIP		
665554	estExt_fgenesh1_pm.C_LG_1_t20245	bZIP		
652780	estExt_Genewise1.C_12_t10209	bZIP		
490310	gw1.1.44.1	bZIP		
667862	estExt_fgenesh1_pm.C_LG_3_t20349	bZIP		IPR021755: Transcription factor Aft1, HRA domain
476130	fgenesh1_kg.12_#_122_#_1315781:2	bZIP		
707132	estExt_fgenesh3_kg.C_LG_9_t20245	bZIP		
636734	estExt_Genewise1Plus.C_12_t20316	C2H2	Crz1	
686144	fgenesh3_kg.LG_3_#_1418_#_Locus_4726_Transcript_2	C2H2		
443509	gm1.7953_g	C2H2	Mig1	
681767	fgenesh3_kg.LG_1_#_2671_#_Locus_5309_Transcript_1	C2H2	Msn4	
296037	Lacbi1.estExt_fgenesh2_pg.C_900005	C2H2	Nrg1	
636637	estExt_Genewise1Plus.C_12_t20172	C2H2	Pzf1	
708164	estExt_fgenesh3_kg.C_11_t10044	C2H2	Rgm1	
379257	Lacbi1.Eugenev2.0000010480	C2H2	Rim101	
479092	fgenesh1_pm.LG_1_#_1809	C2H2	Rpn4	
691497	fgenesh3_kg.LG_7_#_450_#_Locus_1825_Transcript_2/	C2H2	Sfp1	
607158	e_gw1.10.489.1	C2H2	Zap1	
458584	estExt_Genemark1.C_LG_9_t20067	C2H2		
681707	fgenesh3_kg.LG_1_#_2611_#_Locus_459_Transcript_1/	C2H2		
633206	estExt_Genewise1Plus.C_LG_9_t10277	C2H2		
444635	gm1.9079_g	C2H2		
292045	Lacbi1.estExt_fgenesh2_pg.C_20275	C2H2		
705820	estExt_fgenesh3_kg.C_LG_6_t30132	C2H2		
487295	fgenesh1_pm.LG_10_#_915	C2H2		
667917	estExt_fgenesh1_pm.C_LG_3_t20436	C2H2		
301563	Lacbi1.eu2.Lbscf0001g10120	C2H2		
699998	estExt_fgenesh3_kg.C_LG_1_t30319	C2H2		
705977	estExt_fgenesh3_kg.C_LG_7_t10247	C2H2		
453082	gm1.17526_g	C2H2		
480375	fgenesh1_pm.LG_2_#_513	C2H2		
308583	Lacbi1.eu2.Lbscf0003g02220	C2H2		IPR001138: Zn(2)-C6 fungal-type DNA-binding domain

311512	Lacbi1.eu2.Lbscf0004g04680	C2H2		
696532	fgenesh3_kg.12_#_393_#_Locus_5636_Transcript_2/3_	C2H2		
667917	estExt_fgenesh1_pm.C_LG_3_t20436	C2H2		
579433	e_gw1.2.396.1	C2H2		IPR011009: Protein kinase-like domain
670084	estExt_fgenesh1_pm.C_LG_9_t10472	C2H2		IPR001138: Zn(2)-C6 fungal-type DNA-binding domain
621250	e_gw1.25.114.1	C2H2		
297378	Lacbi1.eu2.Lbscf0011g05350	C2H2		
690446	fgenesh3_kg.LG_6_#_618_#_Locus_5398_Transcript_4/	C2H2		
608663	e_gw1.10.4503.1	C2H2		
704012	estExt_fgenesh3_kg.C_LG_3_t50392	C2H2		
579118	e_gw1.2.3753.1	C2H2		
302523	Lacbi1.eu2.Lbscf0021g01280	C2H2		
302525	Lacbi1.eu2.Lbscf0021g01300	C2H2		
643886	estExt_Genewise1.C_LG_3_t110180	C2H2		
445893	gm1.10337_g	C2H2		
709867	estExt_fgenesh3_kg.C_180082	C2H2		
706259	estExt_fgenesh3_kg.C_LG_8_t10085	C2H2		
308011	Lacbi1.eu2.Lbscf0038g00930	C2H2		
617721	e_gw1.16.672.1	C2H2		
310878	Lacbi1.eu2.Lbscf0048g01100	C2H2		
441832	gm1.6276_g	C2H2		
484764	fgenesh1_pm.LG_6_#_532	C2H2		
697356	fgenesh3_kg.14_#_3_#_Locus_2161_Transcript_2/3_Co	C2H2		
467042	fgenesh1_pg.LG_10_#_283	C2H2		
451082	gm1.15526_g	C2H2		
334161	Lacbi1.fgenesh3_pg.C_scaffold_60000009	C2H2		
487565	fgenesh1_pm.11_#_150	C2H2		
445505	gm1.9949_g	C2H2		
623651	estExt_Genewise1Plus.C_LG_1_t170102	C2H2		
150072	Lacbi1.gww1.21.88.1	C2H2		
393192	Lacbi1.Eugenev2.0000360079	C2H2		
619068	e_gw1.19.57.1	Forkhead/Winged helix	Fhl1	IPR000253: Forkhead-associated (FHA) domain; IPR008984: SMAD/FHA domain
699455	estExt_fgenesh3_kg.C_LG_1_t10336	Forkhead/Winged helix	Hsf1	IPR000232: Heat shock factor (HSF)-type, DNA-binding

481652	fgenesh1_pm.LG_3_#_652	Forkhead/Winged helix	Rfx1	
622851	estExt_Genewise1Plus.C_LG_1_t110414	Forkhead/Winged helix		
486945	fgenesh1_pm.LG_10_#_565	Forkhead/Winged helix		
441703	gm1.6147_g	Forkhead/Winged helix		IPR000232: Heat shock factor (HSF)-type, DNA-binding
625683	estExt_Genewise1Plus.C_LG_2_t10329	Forkhead/Winged helix		IPR000232: Heat shock factor (HSF)-type, DNA-binding
164524	Lacbi1.gwh1.30.48.1	Forkhead/Winged helix		IPR011991: Winged helix-turn-helix transcription repressor DNA-binding
607061	e_gw1.10.3220.1	Forkhead/Winged helix		IPR000232: Heat shock factor (HSF)-type, DNA-binding
488576	fgenesh1_pm.12_#_467	GATA	Gat1	IPR013860: Protein of unknown function DUF1752, fungi
614278	e_gw1.13.1484.1	GATA	Gat2	
667316	estExt_fgenesh1_pm.C_LG_3_t10047	GATA		
697879	fgenesh3_kg.16_#_62_#_Locus_1493_Transcript_2/2_C	GATA		
685209	fgenesh3_kg.LG_3_#_483_#_Locus_464_Transcript_5/5	GATA		
293949	Lacbi1.estExt_fgenesh2_pg.C_90336	GATA		IPR013860: Protein of unknown function DUF1752, fungi
306097	Lacbi1.eu2.Lbscf0002g11550	GATA		IPR000014: PAS
668080	estExt_fgenesh1_pm.C_LG_3_t30173	GATA		IPR000014: PAS
585421	e_gw1.3.8714.1	GATA		
671538	estExt_fgenesh1_pm.C_130169	GATA		IPR007205: Uncharacterised protein family UPF0507
617537	e_gw1.16.743.1	GATA		
709764	estExt_fgenesh3_kg.C_160078	GATA		
636228	estExt_Genewise1Plus.C_11_t50003	Helix-turn-helix	Mbf1	IPR010982: Lambda repressor-like, DNA-binding; IPR013729: Multiprotein bridging factor 1, N-terminal
297319	Lacbi1.eu2.Lbscf0011g04760	Homeobox	Pho2	
657699	estExt_fgenesh1_kg.C_LG_4_t10141	Homeobox	Tos8	
379291	Lacbi1.Eugenev2.0000010514	Homeobox		
293988	Lacbi1.estExt_fgenesh2_pg.C_100022	Homeobox		
294716	Lacbi1.estExt_fgenesh2_pg.C_200029	Homeobox		
666027	estExt_fgenesh1_pm.C_LG_1_t30346	Homeobox		
301103	Lacbi1.eu2.Lbscf0001g05520	Homeobox		
308513	Lacbi1.eu2.Lbscf0003g01520	Homeobox		
317188	Lacbi1.eu2.Lbscf0007g03160	Homeobox		
299583	Lacbi1.eu2.Lbscf0017g00370	Homeobox		IPR017442: Serine/threonine-protein kinase-like domain
294716	Lacbi1.estExt_fgenesh2_pg.C_200029	Homeobox		
320896	Lacbi1.fgenesh3_pg.C_scaffold_1000466	Homeobox		

321078	Lacbi1.fgenesh3_pg.C_scaffold_1000648	Homeobox		
327871	Lacbi1.fgenesh3_pg.C_scaffold_14000118	Homeobox		
328498	Lacbi1.fgenesh3_pg.C_scaffold_17000062	Homeobox		
328723	Lacbi1.fgenesh3_pg.C_scaffold_18000055	Homeobox		
399673	Lacbi1.UKU_Eugenev2.0000400105	Homeobox		
336207	Lacbi1.fgenesh3_pg.C_scaffold_409000002	Homeobox		IPR001523: Paired box protein, N-terminal; IPR011991: Winged helix-turn-helix transcription repressor DNA-binding
324166	Lacbi1.fgenesh3_pg.C_scaffold_5000273	Homeobox		
333721	Lacbi1.fgenesh3_pg.C_scaffold_55000004	Homeobox		IPR002622: Transposase, Synechocystis PCC 6803
334913	Lacbi1.fgenesh3_pg.C_scaffold_72000018	Homeobox		
522619	gw1.4.2360.1	MADS box	Mcm1	
302141	Lacbi1.eu2.Lbscf0020g00560	MADS box	Smp1	
293207	Lacbi1.estExt_fgenesh2_pg.C_50417	MADS box		
231949	Lacbi1.e_gwh1.5.218.1	Myb-like factors	Ada2	IPR007526: SWIRM; IPR000433: Zinc finger, ZZ-type
301089	Lacbi1.eu2.Lbscf0001g05380	Myb-like factors	Bdp1	
705566	estExt_fgenesh3_kg.C_LG_6_t20124	Myb-like factors		
190760	Lacbi1.estExt_GeneWisePlus_worm.C_120306	Myb-like factors		
489720	fgenesh1_pm.19_#_117	Myb-like factors		
296675	Lacbi1.eu2.Lbscf0010g04840	Myb-like factors		
705566	estExt_fgenesh3_kg.C_LG_6_t20124	Myb-like factors		IPR021786: Protein of unknown function DUF3351
478107	fgenesh1_pm.LG_1_#_824	SGT1		
628355	estExt_Genewise1Plus.C_LG_3_t40421	Transcription factor, fungi		
690614	fgenesh3_kg.LG_6_#_786_#_Locus_7074_Transcript_2/	YL1 nuclear		
512103	gw1.2.1604.1	zf-NF-X1	Fap1	IPR001374: Single-stranded nucleic acid binding R3H
296682	Lacbi1.eu2.Lbscf0010g04910	*Zinc cluster*	Asg1	
313811	Lacbi1.eu2.Lbscf0005g04000	*Zinc cluster*	Ecm22	
399488	Lacbi1.MEL_Eugenev2.0000050329	*Zinc cluster*	Ert1	
304495	Lacbi1.eu2.Lbscf0028g00440	*Zinc cluster*	Gal4	
459342	estExt_Genemark1.C_11_t10448	*Zinc cluster*	Hap1	
701951	estExt_fgenesh3_kg.C_LG_2_t20198	*Zinc cluster*	Leu3	
573592	e_gw1.1.3345.1	*Zinc cluster*	Oaf3	
670648	estExt_fgenesh1_pm.C_LG_10_t20443	*Zinc cluster*	Rds2	

313099	Lacbi1.eu2.Lbscf0057g00470	*Zinc cluster*		
585149	e_gw1.3.5874.1	*Zinc cluster*	Sef1	
301157	Lacbi1.eu2.Lbscf0001g06060	*Zinc cluster*	Tea1	
629646	estExt_Genewise1Plus.C_LG_3_t90132	*Zinc cluster*	Ume6	
483007	fgenesh1_pm.LG_4_#_118	*Zinc cluster*	Upc2	
381332	Lacbi1.Eugenev2.0000021194	*Zinc cluster*		
670648	estExt_fgenesh1_pm.C_LG_10_t20443	*Zinc cluster*		
292029	Lacbi1.estExt_fgenesh2_pg.C_20246	*Zinc cluster*		
654679	estExt_Genewise1.C_19_t10399	*Zinc cluster*		
708777	estExt_fgenesh3_kg.C_11_t20467	*Zinc cluster*		
625521	estExt_Genewise1Plus.C_LG_1_t90397	*Zinc cluster*		
301029	Lacbi1.eu2.Lbscf0001g04780	*Zinc cluster*		
666247	estExt_fgenesh1_pm.C_LG_1_t40152	*Zinc cluster*		
607166	e_gw1.10.433.1	*Zinc cluster*		
683624	fgenesh3_kg.LG_2_#_551_#_Locus_4406_Transcript_4/7_	*Zinc cluster*		
308722	Lacbi1.eu2.Lbscf0003g03610	*Zinc cluster*		
483154	fgenesh1_pm.LG_4_#_265	*Zinc cluster*		
309166	Lacbi1.eu2.Lbscf0003g08050	*Zinc cluster*		IPR000038: Cell division/GTP binding protein; IPR002409: Aflatoxin biosynthesis regulatory protein
704283	estExt_fgenesh3_kg.C_LG_4_t10098	*Zinc cluster*		
668424	estExt_fgenesh1_pm.C_LG_4_t10027	*Zinc cluster*		
583431	e_gw1.3.717.1	*Zinc cluster*		
685688	fgenesh3_kg.LG_3_#_962_#_Locus_6890_Transcript_4/	*Zinc cluster*		
383085	Lacbi1.Eugenev2.0000040866	*Zinc cluster*		
709278	estExt_fgenesh3_kg.C_12_t20138	*Zinc cluster*		
613245	e_gw1.12.1962.1	*Zinc cluster*		
709256	estExt_fgenesh3_kg.C_12_t20107	*Zinc cluster*		
313796	Lacbi1.eu2.Lbscf0005g03850	*Zinc cluster*		
314000	Lacbi1.eu2.Lbscf0005g05890	*Zinc cluster*		
317008	Lacbi1.eu2.Lbscf0007g01360	*Zinc cluster*		IPR017896: 4Fe-4S ferredoxin, iron-sulpur binding domain
317073	Lacbi1.eu2.Lbscf0007g02010	*Zinc cluster*		
436253	gm1.697_g	*Zinc cluster*		IPR017896: 4Fe-4S ferredoxin, iron-sulpur binding domain
317921	Lacbi1.eu2.Lbscf0008g01030	*Zinc cluster*		
682475	fgenesh3_kg.LG_1_#_3379_#_Locus_3298_Transcript_1	*Zinc cluster*		

707179	estExt_fgenesh3_kg.C_LG_9_t20303	*Zinc cluster*		
623610	estExt_Genewise1Plus.C_LG_1_t160174	*Zinc cluster*		
668161	estExt_fgenesh1_pm.C_LG_3_t30302	*Zinc cluster*		IPR021858: Protein of unknown function DUF3468
486399	fgenesh1_pm.LG_10_#_19	*Zinc cluster*		
579366	e_gw1.2.5378.1	*Zinc cluster*		
300072	Lacbi1.eu2.Lbscf0018g01590	*Zinc cluster*		
593632	e_gw1.4.5086.1	*Zinc cluster*		
614280	e_gw1.13.1325.1	*Zinc cluster*		
697312	fgenesh3_kg.13_#_465_#_Locus_8925_Transcript_1/1_	*Zinc cluster*		
303095	Lacbi1.eu2.Lbscf0023g00790	*Zinc cluster*		
594734	e_gw1.5.2936.1	*Zinc cluster*		
483986	fgenesh1_pm.LG_5_#_125	*Zinc cluster*		
484885	fgenesh1_pm.LG_6_#_653	*Zinc cluster*		
307141	Lacbi1.eu2.Lbscf0034g00060	*Zinc cluster*		
625911	estExt_Genewise1Plus.C_LG_2_t30063	*Zinc cluster*		
595386	e_gw1.5.1735.1	*Zinc cluster*		
310910	Lacbi1.eu2.Lbscf0048g01420	*Zinc cluster*		
312485	Lacbi1.eu2.Lbscf0052g00840	*Zinc cluster*		
315220	Lacbi1.eu2.Lbscf0069g00020	*Zinc cluster*		
486101	fgenesh1_pm.LG_9_#_264	*Zinc cluster*		
654916	estExt_Genewise1.C_220132	*Zinc cluster*		
696110	fgenesh3_kg.11_#_1064_#_Locus_6207_Transcript_3/3	*Zinc cluster*		
567783	e_gw1.1.11887.1	*Zinc cluster*		
299606	Lacbi1.eu2.Lbscf0017g00600	Zinc finger, GRF-type		
625238	estExt_Genewise1Plus.C_LG_1_t90070	Zinc finger, MIZ-type		

Tabella 5: Risultati TAT dei fattori di trascrizione ottenuti dalla genoteche di *Laccaria bicolor* (F:micelio/corpo fruttifero, E:ectomicorriza);

Seq. Name	Gene model	tot	F	E	prot ID	BlastP	organism	score	e-value	Scer	DBD
Contig05	293242	68	58	10	BAC55240.1	C-Gcn4	Candida maltosa	55,5	1E-05	Gcn4	bZIP
ECM-L2_G03	**298274**	**1**	**0**	**1**	**XP_001388805.1**	**regulatory protein abaA**	**Aspergillus niger**	**63,5**	**7E-08**	**Tec1**	**TEA/ATTS**
ECM-L1_F11	307744	1	0	1	XP_002173151.1	meiotically up-regulated gene 152 protein	Schizosaccharomyces japonicus	171	2E-43		Myb
Contig62	386478	7	7	0	XP_001830477.1	PCC1	Coprinopsis cinerea	122	2E-27		HMG1/HMG2
Contig41	393192	3	3	0	XP_001886200.1	STE12-like	Laccaria bicolor	2641	0	Ste12	C2H2
ECM-L1_A03	457991	1	0	1	XP_002910064.1	NWD2	Coprinopsis cinerea	234	2E-61		Forkhead
Contig40	458057	16	15	1	XP_001383328.2	ROX1-like HMG-box transcription factor	Scheffersomyces stipitis	75,5	8E-12		HMG1/HMG2
Contig65	481652	4	3	1	XP_003501812.1	RFX2	Cricetulus griseus	126	3E-27	Rfx1	RFX
FLM-L2_D07	482609	1	1	0	XP_001828950.2	specific transcriptional repressor	Coprinopsis cinerea	109	6E-23	Rox1	HMG1/HMG2
FLM-L2_D05	486090	1	1	0	XP_003034605.1	expressed protein	Schizophyllum commune	168	8E-46		Zn2Cys6
ECM-L2_E05	626440	1	0	1	NP_587684.1	hypothetical protein	Coprinopsis cinerea	337	3E-103		RNA polymerase I-specific transcription initiation factor RRN6-like
Contig46	628355	2	1	1	XP_567555.1	transcriptional regulatory protein	Cryptococcus neoformans	237	6E-63		Transcription factor fungi
FLM-L2_A08	633206	1	1	0	XP_001368548.1	zinc finger protein 850-like	Monodelphis domestica	55,8	2E-05	Met31	C2H2
Contig43	640940	4	3	1	AAC32736.1	Prf1	Ustilago maydis	93,2	2E-17	Rox1	HMG1/HMG2
Contig38	648888	9	6	3	AAC32736.1	Prf1	Ustilago maydis	88,2	2E-15		HMG1/HMG2
ECM-L1_H01	**656449**	**1**	**0**	**1**	**XP_001819986.2**	**regulatory protein abaA**	**Aspergillus oryzae**	**63,9**	**4E-08**		**TEA/ATTS**
Contig75	665554	2	2	0	AAS64313.1	Chap1	Cochliobolus heterostrophus	92.8	3E-17	Yap1	bZIP
Contig07	682475	2	2	0	XP_001399919.1	C6 transcription factor (Mut3)	Aspergillus niger	119	9E-25		Zn2Cys6
Contig42	700637	21	16	5	EGO20236.1	hypothetical protein	Serpula lacrymans	564	0		ZZ Zinc finger

Tabella 6: Risultati TAT delle proteine che non possiedono un DBD condotto sulle genoteche di *Laccaria bicolor* (F: cloni sequenziati della genoteca micelio/corpo fruttifero, E: cloni della genoteca di ectomicorriza). Le righe arancio corrispondono ai cloni che possiedono un dominio di legame ad acidi nucleici.

ANALISI TAT					BLASTP						
Sequ ID	**tot**	**F**	**E**	**gene model**	**acc #**	**BALST result**	**organismo**	**score**	**e-value**	**dominio conservato**	**Sc homolog**
Nuclear proteins and unconventional activators (UAs)											
FLM-L1_F07	1	1	0	192523	XP_001877048.1	tubulin alpha	Laccaria bicolor	932	0		
Contig36	5	5	0	294384	XP_568826.1	tubulin binding protein	Cryptococcus neoformans	286	5E-75		Num1
Contig70	3	0	3	459401	XP_001833620.2	ubiquitin-protein ligase	Coprinopsis cinerea	1461	0		Rsp5
ECM-L2_D08	1	0	1	468224	XP_001828858.2	Rad21 protein	Coprinopsis cinerea	475	2E-158		Mcd1
FLM-L2_D02	1	1	0	700143	XP_001828708.2	CMGC/RCK/MAK protein kinase	Coprinopsis	1117	0		Ime2
FLM-L2_F09	1	1	0	610588	BAG24499.1	rad57	Coprinopsis cinerea	424	7E-143		Rad57
Contig72	2	0	2	636246	CCA72600.1	related to EDE1 protein involved in endocytosis	Piriformospora indica	572	1E-179		Ede1
Contig27	2	0	2	451329	CCA68518.1	related to 40S ribosomal protein	Piriformospora indica	444	9E-134		
ECM-L2_D05	1	0	1	685195	NP_595780.1	ribosome biogenesis protein Nop6	Schizosaccharomyces pombe	83,6	3E-14		
ECM-L1_B01	1	0	1	698517	XP_002472728.1	60S acidic ribosomal protein P1	Postia placenta	125	7E-36		Rpp1
Contig39	2	2	0	585018	XP_001836279.2	hypothetical protein			3E-129	ARID/BRIGHT	
Contig45	2	2	0	699941	XP_001828564.2	hypothetical protein			0	PHD	
FLM-L1_B12	1	1	0	640654	XP_003507462.1	hypothetical protein			7E-51	C3HC4	
ECM-L1_E11	1	0	1	705628	XP_002172787.1	cps3			1E-17	CCCH	Lee1
Putative unconventional activators (PUAs)											
Contig03	3	0	3	190404	XP_001880663.1	aspartic peptidase A1	Laccaria bicolor	944	0		
FLM-L3_C01	1	1	0	299583	XP_001836340.2	TKL/TKL-ccin protein kinase	Coprinopsis	327	6E-95		
ECM-L3_B06	1	0	1	324430	XP_001877591.1	copper transporter	Laccaria bicolor	1432	0		
ECM-L3_A03	1	0	1	327303	XP_001274183.1	mitochondrial GTPase (YlqF)	Aspergillus clavatus	130	7E-32		Mtg1

Contig01	4	4	0	444552	XP_001835217.1	peroxin19	Coprinopsis cinerea	279	6E-89	
ECM-L3_D03	1	0	1	521043	CCA67049.1	related to PDR16-involved in lipid biosynthes	Piriformospora indica	330	7E-111	Pdr16
Contig09	8	8	0	583617	CCA71746.1	related to proteophosphoglycan ppg4	Piriformospora indica	106	2E-20	
ECM-L2_F10	1	0	1	608638	XP_001831367.1	gamma-adaptin	Coprinopsis cinerea okayam	1390	0	Apl4
FLM-L2_E09	1	1	0	671307	XP_001830051.1	peroxisomal targeting signal 1 receptor	Coprinopsis cinerea	771	0	Pex5
FLM-L2_D03	1	1	0	695354	XP_001840019.1	mitochondrial carrier protein	Coprinopsi	476	1E-169	Ggc1
ECM-L2_E11	1	0	1	703237	XP_002911229.1	rho GDP-dissociation inhibitor	Coprinops	330	6E-114	Rdi1p
FLM-L2_A07	1	1	0	707485	XP_002910841.1	guanine nucleotide exchange factor Vps9	Coprinopsis cinerea	937	0	Vps9p
Contig10	2	1	1	509577	XP_001882083.1	predicted protein	Laccaria bicolor	2773	0	
FLM-L2_F03	1	1	0	656382	XP_001875331.1	predicted protein	Laccaria bicolor	600	0	
ECM-L1_B05	1	0	1	311818	XP_003037815.1	hypothetical protein SCHCODRAFT_102409	Schizophyllum commune	55,6	5E-06	
FLM-L2_F11	1	1	0	321043	XP_001829416.1	hypothetical protein CC1G_00595	Coprinopsis cinerea	480	4E-156	
ECM-L2_D01	1	0	1	325350	P_001837732.1	hypothetical protein CC1G_06938	Coprinopsis cinerea	151	4E-34	
FLM-L2_H02	1	1	0	390988	XP_001841219.1	hypothetical protein CC1G_11382	Coprinopsis cinerea	72,8	2E-13	
FLM-L1_C04	1	1	0	439929	EGO18585.1	hypothetical protein	Serpula lacrymans	228	7E-63	
Contig32	3	0	3	459061	XP_003026161.1	hypothetical protein	Schizophyllum commune	61.6	7E-07	
FLM-L3_F02	1	1	0	546684	EGO29111.1	hypothetical protein	Serpula lacrymans	174	3E-50	
Contig54	5	4	1	549772	XP_001833470.2	hypothetical protein CC1G_05170	Coprinopsis cinerea	410	1E-123	
Contig76	5	5	0	576504	XP_001828840.2	hypothetical protein	Coprinopsis cinerea	210	8E-59	
Contig37	2	2	0	604174	XP_001834463.1	hypothetical protein CC1G_02199	Coprinopsis cinerea	541	3E-169	
FLM-L1_G10	1	1	0	634434	XP_001831375.1	hypothetical protein CC1G_00922	Coprinopsis cinerea	373	4E-117	
FLM-L2_E07	1	1	0	686252	XP_001836432.2	hypothetical protein CC1G_07079	Coprinopsis cinerea	258	1E-71	
ECM-L1_B02	1	0	1	693899	EGO22612.1	hypothetical protein	Serpula lacrymans	336	1E-105	
FLM-L1_E04	1	1	0	708222	XP_003036324.1	expressed protein	Schizophyllum commune	87	3E-15	
Contig68	2	0	2	708574	XP_001833764.1	hypothetical protein	Coprinopsis cinerea	330	5E-104	
Contig24	2	2	0	613652	XP_001830379.1	hypothetical protein	Coprinopsis cinerea	180	2E-43	

Contig78	5	5	0	680010	XP_002476516.1	predicted protein	Postia placenta	114	2E-24	
Secreted protein										
FLM-L2_D04	1	1	0	243371	XP_002912138.1	aconitate hydratase	Coprinopsis cinerea	1244	0	Aco2p
FLM-L2_B06	1	1	0	300643	CCA73543.1	related to meiotic recombination protein rec8	Piriformospora indica	122	1E-26	Rec8p
ECM-L1_H02	1	0	1	394934	ZP_01463678.1	sphingolipid ceramide N-deacylase	Stigmatella aurantiaca	62,4	2E-08	
ECM-L3_B04	1	0	1	622202	XP_001828856.1	hypothetical protein CC1G_03650	Coprinopsis cinerea	662	0	
Contig44	3	3	0	643792	XP_001836617.1	hypothetical protein CC1G_06204	Coprinopsis cinerea	414	3E-128	
FLM-L2_H04	1	1	0	656953	XP_001837070.1	vacuole protein	Coprinopsis cinerea okay	579	0	YJR054Wp-like protein
FLM-L3_H05	1	1	0	658920	XP_001830283.1	hypothetical protein CC1G_01919	Coprinopsis cinerea	124	6E-30	
FLM-L2_G08	1	1	0	660445	XP_001835021.1	hypothetical protein CC1G_09912	Coprinopsis cinerea	221	1E-69	
FLM-L1_G04	1	1	0	688063	XP_001835466.2	hypothetical protein CC1G_05428	Coprinopsis cinerea	160	2E-42	
ECM-L3_C02	1	0	1	688275	XP_001835642.2	hypothetical protein CC1G_03424	Coprinopsis cinerea	118	5E-27	
Contig58	4	0	4	693322	XP_001830639.2	hypothetical protein CC1G_06905	Coprinopsis cinerea	1292	0	
Secreted proteins with nuclear localization										
FLM-L1_A04	1	1	0	304792	XP_001840014.1	hypothetical protein CC1G_10398	Coprinopsis cinerea	60,1	9E-08	
FLM-L1_F01	1	1	0	391051	XP_001840014.1	hypothetical protein	Coprinopsis cinerea	60,5	2E-07	
FLM-L1_D07	1	1	0	455116	XP_001884865.1	predicted protein	Laccaria bicolor	664	0	
Contig64	8	7	1	659547	AAD01986.1	ras related protein	Laccaria bicolor	200	2E-61	
FLM-L2_E08	1	1	0	659644	XP_001839900.1	elongation factor 3	Coprinopsis cinerea	1819	0	Hef3p

Tabella 7: Fattori di trascrizione di pianta isolati mediante *screening* TAT delle genoteche di ectomicorriza di *Laccaria bicolor* e di radice di *Populus trichocarpa* (E: cloni sequenziati della genoteca di ectomicorriza, R: cloni della genoteca di pioppo)

Seq. Name	tot	E	R	**prot ID**	**description**	**organism**	**score**	**e-value**
Contig04	23	10	13	XP_002304249.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	780	0
Contig06	4	1	3	ACV92010.1	WRKY transcription factor 8	Populus tomentosa	502	4E-180
Contig08	3	1	2	XP_002518861.1	transcription factor, putative	Ricinus communis	1397	0
Contig11	46	17	29	XP_002328610.1	type-b response regulator	Populus trichocarpa	1390	0
Contig14	3	0	3	XP_003632476.1	homeobox-leucine zipper protein	Vitis vinifera	429	5E-149
Contig15	2	2	0	XP_002312841.1	ethylene-insensitive 3a	Populus trichocarpa	1316	0
Contig16	2	1	1	AAS00054.1	CONSTANS-like protein CO1	Populus deltoides	700	0
Contig18	2	2	0	XP_002524697.1	r2r3-myb transcription factor	Ricinus communis	426	3E-147
Contig19	3	3	0	ACF77118.1	CBF4a	Populus hopeiensis	464	1E-165
Contig20	2	2	0	XP_002301042.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	669	0
Contig22	2	2	0	XP_002310688.1	NAC domain protein	Populus trichocarpa	640	0
Contig25	12	7	5	XP_002331217.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	774	0
Contig26	6	3	3	XP_002531120.1	r2r3-myb transcription factor	Ricinus communis	372	9E-130
Contig28	4	2	2	XP_002311456.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	640	0
Contig29	3	3	0	XP_002315338.1	NAC domain protein	Populus trichocarpa	1137	0
Contig30	2	2	0	XP_002315400.1	ethylene-insensitive 3c	Populus trichocarpa	1263	0
Contig31	3	3	0	XP_002307195.1	NAC domain protein	Populus trichocarpa	639	0

Contig34	2	2	0	XP_002319156.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	562	0
Contig35	5	3	2	XP_002298067.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	473	1E-169
Contig47	7	2	5	XP_002306731.1	NAC domain protein	Populus trichocarpa	1192	0
Contig48	2	1	1	XP_002511920.1	WRKY transcription factor	Ricinus communis	398	6E-137
Contig49	8	3	5	ADW65758.1	CONSTANS-like protein	Gossypium hirsutum	467	7E-164
Contig51	3	1	2	XP_002302150.1	NAC domain protein	Populus trichocarpa	1276	0
Contig55	4	2	2	XP_002318846.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	395	3E-139
Contig56	2	0	2	ADL36694.1	GATA domain class transcription factor	Malus x domestica	308	2E-101
Contig57	11	6	5	XP_002528697.1	WRKY transcription factor	Ricinus communis	300	1E-98
Contig61	4	1	3	XP_002266803.1	transcription factor VIP1-like	Vitis vinifera	390	2E-133
Contig63	3	1	2	ACV92010.1	WRKY transcription factor 8	Populus tomentosa	438	9E-155
Contig66	2	0	2	XP_002274709.1	RING finger and CHY zinc finger domain-containing protein	Vitis vinifera	483	9E-172
Contig67	10	1	9	BAG15872.1	transcription factor	Bruguiera gymnorhiza	880	0
Contig69	2	1	1	XP_002310574.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	775	0
Contig73	2	1	1	XP_002534043.1	GATA transcription factor	Ricinus communis	434	1E-150
Contig74	2	1	1	XP_003525349.1	probable WRKY transcription factor	Glycine max	402	6E-138
Contig79	2	0	2	XP_002515013.1	r2r3-myb transcription factor	Ricinus communis	233	1E-77
Contig80	2	0	2	XP_002533810.1	nuclear receptor binding set domain containing protein 1, nsd	Ricinus communis	1258	0

Contig81	2	0	2	XP_002533687.1	zinc finger protein	Ricinus communis	471	2E-164
Contig82	2	2	0	NP_566998.1	Forkhead-associated (FHA) domain	Arabidopsis thaliana	215	1E-59
ECM-L1_A05	1	1	0	XP_002311983.1	NAC domain protein	Populus trichocarpa	1177	0
ECM-L1_C02	1	1	0	XP_002323615.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	972	0
ECM-L1_D10	1	1	0	XP_002519505.1	DNA binding protein	Ricinus communis	983	0
ECM-L1_E05	1	1	0	XP_002300526.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	699	0
ECM-L1_E09	1	1	0	XP_002326913.1	f-box family protein	Populus trichocarpa	537	0
ECM-L2_B02	1	1	0	XP_003544908.1	probable WRKY transcription factor	Glycine max	707	0
ECM-L2_C12	1	1	0	XP_002308444.1	transcription factor E2F	Populus trichocarpa	875	0
ECM-L2_F01	1	1	0	XP_002266803.1	transcription factor VIP1-like	Vitis vinifera	332	1E-110
ECM-L3_B05	1	1	0	XP_002332559.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	707	0
ECM-L3_G04	1	1	0	XP_002328098.1	ethylene-insensitive 3b	Populus trichocarpa	1293	0
ROOT-L1_A03	1	0	1	XP_002309945.1	NAC domain protein	Populus trichocarpa	638	0
ROOT-L1_C05	1	0	1	NP_201474.1	C2H2-like zinc finger protein	Arabidopsis thaliana	518	0
ROOT-L1_E05	1	0	1	XP_002512054.1	LIGULELESS1 protein, putative	Ricinus communis	239	5E-74
ROOT-L1_G09	1	0	1	XP_003535147.1	transcription factor GAMYB-like	Glycine max	347	7E-117
ROOT-L2_A07	1	0	1	XP_003593218.1	Methyl binding domain protein	Medicago truncatula	479	6E-155
ROOT-L2_C07	1	0	1	XP_002334238.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	559	0
ROOT-L2_C09	1	0	1	XP_002307390.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	514	0
ROOT-L2_D09	1	0	1	XP_002301490.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	692	0

ROOT-L2_E07	1	0	1	ADW65758.1	CONSTANS-like protein	Gossypium hirsutum	473	5E-166
ROOT-L2_E11	1	0	1	XP_002313234.1	GRAS family transcription factor	Populus trichocarpa	1113	0
ROOT-L2_H05	1	0	1	XP_002313601.1	NAC domain protein, IPR003441	Populus trichocarpa	647	0
ROOT-L3_B12	1	0	1	XP_003604938.1	Auxin response factor	Medicago truncatula	776	0
ROOT-L3_E12	1	0	1	AAV85474.1	myb family transcription factor-related protein	Populus tomentosa	598	0
ROOT-L3_G11	1	0	1	XP_002318847.1	AP2/ERF domain-containing transcription factor	Populus trichocarpa	558	0

Tabella 8: Proteine prive di un DBD isolate dallo *screening* TAT in pianta (all'interno delle proteine nucleari in grassetto sono evidenziate le sequenze che possiedono domini di legame agli acidi nucleici che non sono caratterizzati come domini di legame al DNA propri di fattori di trascrizione)

Nuclear proteins and unconventional activators (UAs)								
Seq. Name	**tot**	**ECM**	**ROOT**	**prot ID**	**description**	**organism**	**score**	**e-value**
Contig02	5	1	4	XP_002533858.1	26S proteasome non-atpase regulatory subunit	Ricinus communis	563	0
Contig17	2	1	1	NP_196773.1	Ubiquitin-associated/translation elongation factor EF1B	Arabidopsis thaliana	268	4,00E-85
Contig21	3	3	0	NP_193405.1	60S ribosomal protein L15-1	Arabidopsis thaliana	349	2,00E-124
Contig33	2	1	1	AAF34804.1	CDK-activating kinase	Euphorbia esula	699	0
Contig50	2	0	2	NP_180260.1	Ubiquitin-associated/translation elongation factor EF1B	Arabidopsis thaliana	505	3,00E-169
Contig52	2	0	2	XP_002280700.1	protein LTV1 homolog	Vitis vinifera	506	3,00E-173
Contig59	9	7	2	XP_002275593.1	importin subunit alpha-1	Vitis vinifera	1004	0
Contig60	2	0	2	XP_003595566.1	CTD small phosphatase-like protein	Medicago truncatula	533	0
Contig83	2	1	1	XP_002528348.1	o-linked n-acetylglucosamine transferase	Ricinus communis	286	8,00E-93
ECM-L1_B08	1	1	0	XP_002888582.1	proteasome maturation factor UMP1 family	Arabidopsis lyrata	192	1,00E-61
ECM-L1_G09	1	1	0	XP_002514774.1	cohesin subunit rad21	Ricinus communis	1264	0
ECM-L1_H10	1	1	0	XP_002283831.2	26S protease regulatory subunit 6A homolog A	Vitis vinifera	840	0
ECM-L2_B04	1	1	0	XP_002271062.1	E3 ubiquitin-protein ligase RING1-like	Vitis vinifera	140	4,00E-38
ECM-L2_D07	1	1	0	XP_002532142.1	transcription elongation factor s-II	Ricinus communis	781	0
ROOT-L1_E11	1	0	1	XP_002532142.1	transcription elongation factor s-II	Ricinus communis	663	0
ECM-L2_H11	1	1	0	XP_002510837.1	Transcriptional corepressor SEUSS	Ricinus communis	1018	0
ROOT-L1_C04	1	0	1	NP_565199.1	DNA topoisomerase 2-associated protein PAT1	Arabidopsis thaliana	800	0
ROOT-L1_F06	1	0	1	XP_002284945.1	E3 SUMO-protein ligase SIZ1	Vitis vinifera	767	0
ROOT-L1_F10	1	0	1	BAF46306.1	ataxin-2 related protein	Ipomoea nil	493	4,00E-167
ROOT-L1_G07	1	0	1	XP_002521404.1	DNA repair helicase rad5,16,	Ricinus communis	91,3	1,00E-18
ROOT-L2_G05	1	0	1	XP_003630417.1	Serine/threonine protein kinase	Medicago truncatula	626	0

ROOT-L1_D06	1	0	1	XP_002523135.1	kinesin heavy chain	Ricinus communis	1737	0
ECM-L2_F03	1	1	0	XP_002511999.1	**RNA binding protein**	Ricinus communis	1238	0
ROOT-L2_E05	1	0	1	XP_002279792.2	**protein MEI2-like 2-like**	Vitis vinifera	1198	0
ECM-L1_E02	1	1	0	XP_002279202.2	**zinc finger CCCH domain-containing protein**	Vitis vinifera	1000	0
ROOT-L2_D04	1	0	1	XP_002268751.1	**pumilio homolog 5**	Vitis vinifera	840	0
Putative unconventional activators (PUAs)								
Seq. Name	**tot**	**ECM**	**ROOT**	**prot ID**	**description**	**organism**	**score**	**e-value**
Contig12	6	1	5	XP_002514982.1	3'-N-debenzoyl-2'-deoxytaxol N-benzoyltransferase	Ricinus communis	533	0
Contig23	2	1	1	NP_179621.1	RPM1 interacting protein 13	Arabidopsis thaliana	88,2	6,00E-16
Contig71	3	1	2	ABJ91226.1	CBL-interacting protein kinase 19	Populus trichocarpa	1077	0
Contig77	2	2	0	XP_002513820.1	Oxysterol-binding protein	Ricinus communis	622	0
ECM-L1_A04	1	1	0	XP_002533958.1	SGT1 protein, putative	Ricinus communis	587	0
ECM-L1_B04	1	1	0	XP_002509715.1	ara4-interacting protein	Ricinus communis	746	0
ECM-L1_G10	1	1	0	XP_002513634.1	NADH dehydrogenase	Ricinus communis	292	2,00E-100
ECM-L2_A05	1	1	0	NP_200436.1	protein hapless 8	Arabidopsis thaliana	195	3,00E-51
ECM-L2_B06	1	1	0	XP_002525994.1	Polcalcin Jun	Ricinus commun	246	6,00E-82
ECM-L2_B07	1	1	0	XP_002326245.1	serine/threonine protein kinase 1, CTR1	Populus trichocarpa	1697	0
ECM-L2_C01	1	1	0	XP_002330269.1	2-oxoglutarate-dependent dioxygenase	Populus trichocarpa	759	0
ECM-L2_C08	1	1	0	XP_002893466.1	GYF domain-containing protein	Arabidopsis thaliana	131	8,00E-28
ECM-L2_C11	1	1	0	NP_563791.1	putative protein phosphatase 2C 4	Arabidopsis thaliana	948	0
ECM-L2_D11	1	1	0	XP_002511206.1	Polcalcin	Ricinus communis	194	1,00E-62
ECM-L2_G05	1	1	0	XP_002275793.1	ganglioside-induced differentiation-associated-protein 2	Vitis vinifera	421	1,00E-148
ECM-L2_G07	1	1	0	XP_002533887.1	organic anion transporter, putative	Ricinus communis	617	0
ECM-L3_C03	1	1	0	NP_569023.1	GTP binding protein	Arabidopsis thaliana	419	1,00E-145
ECM-L3_D02	1	1	0	NP_201222.1	dentin sialophosphoprotein-like protein	Arabidopsis thaliana	317	8,00E-99
ROOT-L1_A10	1	0	0	ABK92631.1	ATP binding protein	Ricinus communis	132	8,00E-35
ROOT-L1_B01	1	0	1	XP_002322910.1	aquaporin, MIP family, TIP subfamily	Populus trichocarpa	494	3,00E-177

ROOT-L1_C07	1	0	1	NP_177487.2	protein kinase domain-containing protein	Arabidopsis thaliana	1385	0
ROOT-L1_D10	1	0	1	XP_002529225.1	Cysteine protease ATG4B, putative	Ricinus communis	685	0
ROOT-L1_E08	1	0	1	ACN76468.1	CBL-interacting protein kinase 10	Populus euphratica	790	0
ROOT-L1_E09	1	0	1	XP_002510470.1	vacuolar proton atpase, putative	Ricinus communis	1395	0
ROOT-L1_E10	1	0	1	NP_188412.2	protein kinase family protein	Arabidopsis thaliana	1645	0
ROOT-L1_H03	1	0	1	XP_002285194.1	serine/threonine-protein phosphatase	Glycine max	1343	0
ROOT-L1_H08	1	0	1	NP_564664.1	autophagy 18H-like protein	Arabidopsis thaliana	989	0
ROOT-L1_H11	1	0	1	NP_173582.2	calcium-binding EF-hand-containing protein	Arabidopsis thaliana	855	0
ROOT-L2_B04	1	0	1	XP_003634651.1	receptor-like protein kinase THESEUS 1	Arabidopsis thaliana	1402	0
ROOT-L2_B07	1	0	1	XP_002885828.1	SH3 domain-containing protein	Arabidopsis thaliana	909	0
ROOT-L2_D08	1	0	1	XP_002512737.1	Chaperone protein dnaJ 8, chloroplast precursor	Ricinus communis	264	3,00E-89
ROOT-L2_D10	1	0	1	XP_002318921.1	stress-induced hydrophobic peptide	Populus trichocarpa	105	9,00E-30
ROOT-L2_E01	1	0	1	XP_002510170.1	cysteine protease, putative	Ricinus communis	786	0
ROOT-L2_E02	1	0	1	XP_002516577.1	AP-2 complex subunit beta-1, putative	Ricinus communis	1451	0
ROOT-L2_E03	1	0	1	XP_003632664.1	probable proteasome inhibitor	Vitis vinifera	348	4,00E-118
ROOT-L2_E09	1	0	1	XP_002516207.1	phosphoinositide 5-phosphatase	Ricinus communis	1300	0
ROOT-L2_F07	1	0	1	XP_002516953.1	epoxide hydrolase	Ricinus communis	261	4,00E-86
ROOT-L2_F08	1	0	1	XP_003636025.1	Photosystem II CP47 chlorophyll apoprotein	Medicago truncatula	588	0
ROOT-L2_H03	1	0	1	XP_002324039.1	f-box family protein	Populus trichocarpa	1032	0
ECM-L2_C02	1	1	0	BAC57955.1	polyubiquitin	Aster tripolium	453	4,00E-162
ROOT-L2_H04	1	0	1	XP_002869098.1	cyclase family protein	Arabidopsis lyrat	370	7,00E-128
Contig13	5	3	2	XP_002517818.1	hypothetical protein	Ricinus communis	720	0
Contig53	2	1	1	NP_196277.2	uncharacterized protein	Arabidopsis thaliana	208	8,00E-67
ECM-L1_B06	1	1	0	NP_566370.1	uncharacterized protein	Arabidopsis thaliana	420	2,00E-148
ECM-L1_C05	1	1	0	XP_002526924.1	conserved hypothetical protein	Ricinus communis	1153	0
ECM-L1_D07	1	1	0	ABK94288.1	unknown protein	Populus trichocarpa	132	3,00E-39
ECM-L3_F04	1	1	0	NP_001067108.2	hypothetical protein	Ricinus communis	292	5,00E-92

ROOT-L1_C03	1	0	1	ABK96357.1	hypothetical protein	Populus trichocarpa	489	4,00E-176
ROOT-L1_E06	1	0	1	ACU24189.1	unknown protein	Glycine max	419	3,00E-144
ROOT-L1_F08	1	0	1	BAJ53105.1	predicted protein	Populus trichocarpa	585	0
ROOT-L2_D03	1	0	1	YP_004532427.1	predicted protein	Populus trichocarpa	505	0
ROOT-L2_D07	1	0	1	NP_001043788.1	predicted protein	Populus trichocarpa	1891	0
ROOT-L2_G08	1	0	1	AAL87250.1	unknown protein	Arabidopsis thaliana	496	6,00E-157