

UNIVERSITÀ DEGLI STUDI DI MILANO - BICOCCA
Facoltà di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea Specialistica in Biologia



LA TEORIA DELLE RETI APPLICATA ALLO
STUDIO DELLE SIMBIOSI

Relatore:

Dott. Maurizio CASIRAGHI

Tesi di Laurea di:

Alice DOSE

Matr. 041330

Anno Accademico 2006/2007

INDICE

INTRODUZIONE.....	3
1. LA TEORIA DELLE RETI.....	4
1. Cenni storici.....	5
1.1. La teoria dei grafi.....	5
1.2. La teoria dei grafi casuali.....	6
1.3. La proprietà di piccolo mondo.....	7
1.4. Il clustering.....	9
1.5. La distribuzione dei gradi dei nodi secondo una legge di potenza.....	11
1.6. Nuovi modelli di rete.....	12
2. Modelli di rete.....	12
2.1. Il modello dei grafi casuali.....	12
2.2. Il modello delle reti piccolo-mondo.....	14
2.3. Il modello delle reti a invarianza di scala.....	17
3. Sintesi.....	21
2. LE RETI CELLULARI.....	24
1. Le reti metaboliche.....	24
2. Le reti di interazione tra proteine.....	28
3. GLI ORGANISMI COINVOLTI NELL'ANALISI E IL CONCETTO DI SIMBIOSI.....	31
1. Concetto di simbiosi.....	31
1.1. L'endosimbiosi o simbiosi intracellulare.....	33
2. Gli organismi analizzati.....	34
2.1. Wolbachia.....	34
2.2. Rickettsiales.....	39
2.3. Endosimbionti mutualisti di insetti.....	43
2.4. Batteri a vita libera.....	46
4. METODI.....	49

1. Costruzione delle reti metaboliche.....	49
1.1. Il database.....	49
1.2. Costruzione della rete metabolica.....	49
2. Analisi delle reti metaboliche.....	50
2.1. Distribuzione dei gradi dei nodi.....	50
2.2. Diametro e cammino medio tra i nodi.....	51
2.3. Hub.....	51
2.4. Effetti di attacchi ed errori sulla rete.....	51
3. Analisi delle interazioni con l'ospite.....	52
4. Costruzione dell'albero filogenetico.....	52
5. RISULTATI E DISCUSSIONE.....	53
1. Analisi della rete metabolica di Wolbachia.....	53
1.1. Caratteristiche strutturali della rete.....	53
1.2. Analisi della tolleranza ad attacchi ed errori.....	56
1.3. Analisi delle interazioni con l'ospite.....	60
2. Analisi e confronto delle reti metaboliche.....	63
2.1. Relazioni filogenetiche.....	63
2.2. Analisi delle reti metaboliche.....	64
CONCLUSIONI.....	74
BIBLIOGRAFIA.....	75
APPENDICI.....	81

INTRODUZIONE

Lo scopo di questo lavoro è quello di studiare le reti metaboliche di diversi microrganismi simbiotici mediante l'applicazione della teoria delle reti. Questa disciplina, nata negli ultimi decenni, ha come obiettivo la rappresentazione dei sistemi reali mediante delle reti, cioè degli insiemi di nodi connessi da archi. La realtà così modellizzata viene quindi studiata a partire dalle proprietà strutturali e funzionali di queste reti.

In una rete metabolica i nodi sono rappresentati dai substrati e prodotti, mentre le reazioni costituiscono i collegamenti tra questi nodi. Per la presente indagine le reti sono state ricostruite a partire dai pathway metabolici degli organismi di interesse, per i quali fossero disponibili dati derivanti da progetti di sequenziamento genomico. In particolare si è partiti dall'analisi della rete metabolica di *Wolbachia pipientis*, simbionte intracellulare del nematode filaride *Brugia malayi*. Questa è stata poi confrontata con le reti metaboliche di altri simbiotici filogeneticamente affini a *Wolbachia*, di endosimbionti di insetti (*Wigglesworthia glossinidia*, *Sodalis glossinidius* e *Buchnera aphidicola*), oltre alle reti di batteri a vita libera introdotte come paragone (*Escherichia coli*, *Agrobacterium tumefaciens*, *Salmonella enterica* e *Magnetospirillum magneticum*). Lo scopo di questi confronti è quello di verificare la presenza di differenze nelle reti metaboliche ricostruite legate allo stile di vita o alle relazioni filogenetiche tra questi batteri.

Di ogni rete sono stati studiati gli aspetti strutturali, al fine di ricondurre la loro architettura ad uno dei modelli proposti dalla teoria delle reti. Attraverso questa analisi è stato possibile rilevare alcune differenze nelle reti metaboliche di organismi simbiotici e a vita libera, e individuare le molecole chiave del metabolismo cellulare. Attraverso la teoria delle reti viene quindi proposto un metodo per studiare le interazioni tra gli organismi coinvolti in una simbiosi individuando i possibili scambi di sostanze che avvengono tra simbiotici ed ospiti.

Sono attualmente in corso progetti di sequenziamento dei genomi di *Wolbachia*, e in particolare, per uno di nematodi (*Brugia malayi*) il laboratorio presso cui ho svolto la tesi ha avuto delle collaborazioni con il consorzio internazionale. La mia tesi rappresenta quindi un tentativo per una innovativa analisi genomico-comparativa.

1. La teoria delle reti

La teoria delle reti è una disciplina relativamente nuova, sviluppatasi negli ultimi decenni. Essa ha come oggetto di studio i sistemi complessi, che vengono rappresentati come un insieme di nodi collegati tra loro a formare una rete. Qualsiasi sistema reale, naturale o sociale, può essere rappresentato in questo modo: i rapporti interpersonali nella rete sociale, le interazioni tra molecole nella cellula, le reti ecologiche, internet, la rete neurale e molte altre. Come affermano Newman, Barabási e Watts (2006), i principali studiosi di questo campo, “le reti sono ovunque”. Studiando i sistemi complessi da questa prospettiva, si può avere una maggiore comprensione della loro struttura, funzionamento e comportamento. La teoria delle reti si pone come superamento del riduzionismo, che in epoche precedenti aveva portato a scomporre i sistemi nelle loro componenti elementari, che venivano studiate singolarmente. Una volta descritte tutte le parti, ci si aspettava, assemblandole, di poter comprendere il tutto: in realtà questo non è possibile, in quanto la complessità dei sistemi reali non è riconducibile alla semplice somma delle loro parti. Si è capito, infatti, che la complessità deriva dal fatto che i costituenti dei sistemi reali interagiscono e si uniscono in modi molto diversi. Attraverso la prospettiva delle reti si vuole quindi riassemblare queste parti ricostruendo le mappe delle loro interazioni, in modo da comprendere i sistemi reali nella loro totalità. Per fare questo, dapprima si studia l’architettura di base dei sistemi, osservando solo i nodi e i link che formano la rete, per poi considerare le dinamiche che hanno luogo nei collegamenti, andando oltre la semplice struttura (Barabási, 2004).

Negli ultimi anni la teoria delle reti è stata applicata in molti ambiti, anche molto diversi tra loro, quali la sociologia, la fisica, l’informatica, l’ecologia, la biologia, l’economia. Questo dimostra l’importanza che sta assumendo la nuova scienza delle reti per la comprensione del mondo che ci circonda, tanto da essere definita la “prossima rivoluzione scientifica” (Barabási, 2004).

Il presente capitolo, in cui verrà esposta la teoria delle reti, sarà suddiviso in tre sezioni:

- nella prima parte verranno descritti la nascita e lo sviluppo della teoria dal punto di vista storico;

- nella seconda parte verranno presentati i modelli matematici utilizzati per la descrizione delle reti;
- nella terza parte saranno riassunti i punti principali della teoria, per fornire una più rapida chiave di lettura al mio lavoro di tesi.

1. CENNI STORICI

Lo studio delle reti ha una lunga storia, con radici nella matematica, a partire dalla teoria dei grafi.

1.1. La teoria dei grafi

La teoria dei grafi è un ramo della matematica che studia i diversi modi in cui determinati oggetti si connettono. Per grafo si intende un insieme di vertici, o nodi, connessi da archi, o link.

Si è soliti far risalire l'origine della teoria dei grafi al 1736, con Leonard Eulero e la sua risoluzione del cosiddetto problema dei sette ponti di Königsber. Questa cittadina nei pressi di San Pietroburgo, si trovava sul fiume Pregel, sul quale erano stati costruiti sette ponti: il problema consisteva nel trovare un percorso intorno alla città che permettesse di attraversare ciascun ponte una sola volta (Figura 1). Eulero risolse il problema dimostrando matematicamente l'impossibilità di una sua soluzione, con quello che venne poi considerato da molti il primo teorema della teoria dei grafi: su un grafo che abbia più di due nodi con un numero dispari di link un percorso che attraversi tutti i nodi una sola volta non può esistere.

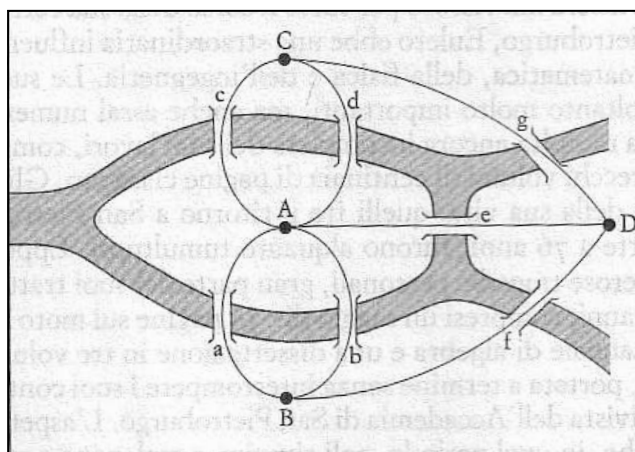


Fig. 1. La pianta schematica di Königsberg, che illustra il problema dei sette ponti (da Barabási, 2004).

Questa dimostrazione di Eulero è importante in quanto egli ha concepito i ponti come un grafo. Da questo momento si comprende, quindi, che la realtà può essere non solo rappresentata da grafi, o reti, ma anche essere spiegata attraverso le proprietà di questi grafi. Le reti possono perciò essere usate come modello per la comprensione dei sistemi reali.

Dopo Eulero, la teoria dei grafi ha avuto un grande sviluppo, vertendo soprattutto sullo studio dei grafi regolari, cioè grafi in cui tutti i nodi hanno lo stesso numero di collegamenti, come ad esempio il reticolo degli atomi in un cristallo o quello delle cellette esagonali nei favi delle api.

Fino alla metà del XX secolo l'obiettivo principale di questa teoria è stato quello di catalogare le proprietà dei vari grafi. Solo a partire dagli anni '50 del secolo scorso, con gli studi di Erdős e Rényi sui grafi casuali, si passò dallo studio delle proprietà dei grafi a quello della loro nascita e formazione.

1.2. La teoria dei grafi casuali

La teoria dei grafi casuali si basa sul principio che la probabilità di connessione è uguale per qualsiasi coppia di nodi della rete, e si distribuisce in modo casuale. Le reti nascono quindi scegliendo due nodi a caso e collegandoli. Se a questo punto si aggiunge un numero di link tale per cui ogni nodo ne abbia almeno uno, si ottiene complessivamente un unico gruppo di nodi interamente connesso. Infatti le reti casuali sono molto connesse, e caratterizzate da un basso diametro, valore che rappresenta una misura della distanza massima tra ogni coppia di nodi, espressa dal numero di link necessari per unire i due nodi. Possiedono cioè la proprietà che è stata poi definita di "piccolo mondo": partendo da un nodo qualunque, è possibile raggiungere qualsiasi altro nodo della rete in pochi passaggi. Questa particolarità è stata descritta per la prima volta nella rete sociale, e da qui ha avuto origine la famosa espressione i "sei gradi di separazione". Si è visto, infatti, che nella rete sociale mondiale umana sono sufficienti in media sei passaggi (link rappresentati da contatti/ conoscenti tra le persone) per connettere due individui qualsiasi. Dal punto di vista della teoria dei grafi, questa espressione si può tradurre nel modo seguente: la società umana è una rete di 7 miliardi di nodi, dove la distanza media fra due nodi qualsiasi è di sei link.

1.3. La proprietà di piccolo mondo

Il concetto dei sei gradi di separazione ha un'origine indipendente rispetto alla teoria dei grafi. La prima apparizione nota del concetto si trova in un'antologia di racconti ("Ogni cosa è diversa") scritta da Frigyes Karinthy, un autore ungherese, nel 1929. In uno di questi racconti, intitolato "Catene", un personaggio propone agli altri un esperimento, nel quale ognuno avrebbe dovuto scegliere una persona qualsiasi nel mondo, a cui avrebbe potuto collegarsi attraverso una catena di sole cinque persone. Sebbene introduce il concetto, nel racconto Karinthy non parla di "sei gradi di separazione". L'origine di questa espressione è da attribuire a John Guare. In una sua commedia del 1991 si legge: "Ho letto che ognuno di noi su questo pianeta è separato dagli altri solo da sei persone. Sei gradi di separazione tra noi e tutti gli altri su questo pianeta...".

Quello a cui John Guare si riferiva era probabilmente uno tra i primi studi quantitativi sul fenomeno, condotto da Stanley Milgram (1933-1984), psicologo americano nel 1967. Lo scopo dell'esperimento era quello di misurare la "distanza sociale" tra due cittadini qualsiasi degli Stati Uniti (Travers e Milgram, 1969). Milgram scelse casualmente alcune persone in Kansas e in Nebraska, e chiese loro di consegnare una lettera ad un operatore di borsa a Boston, di cui era noto solo il nome. Dovevano svolgere questo compito non inviando la lettera all'indirizzo del destinatario, ma consegnandola a conoscenti diretti, che ritenevano avessero qualche punto di contatto con esso. Milgram calcolò il numero di passaggi di ogni lettera che aveva raggiunto il destinatario: in media erano stati necessari 5,5 intermediari perché la lettera arrivasse all'operatore di borsa. Da qui l'idea che i sei gradi di separazione tra due individui di una società altamente connessa come quella americana fossero sei. In realtà i risultati di questo esperimento furono soggetti a critiche: molte lettere non arrivarono al destinatario, e quindi quelle arrivate non davano un quadro fedele della situazione. Le lettere mai arrivate potrebbero aver seguito un percorso più lungo di sei passaggi, prima di essere buttate. Il numero dei gradi di separazione sarebbe quindi stato sottostimato (Buchanan, 2003). Ma altri studi sottolineano come questo valore avrebbe potuto anche essere stato sovrastimato (Barabási, 2004). Infatti tra due persone esistono molti percorsi, di lunghezza diversa: ogni persona a seconda delle disposizioni e delle preferenze, può aver deciso di far pervenire la lettera attraverso certe persone e non altre, scegliendo un percorso che non è necessaria-

mente quello più breve. Infine, l'esperimento riguardava soltanto gli Stati Uniti: i risultati potevano essere generalizzati?

Studi successivi, riguardanti altre reti sociali, come la rete degli attori di Hollywood e la rete delle citazioni negli studi scientifici hanno portato a rispondere positivamente a questa domanda: tutte le reti sociali hanno proprietà di piccolo-mondo.

La dimostrazione che la rete degli attori è un piccolo mondo la si può avere dal cosiddetto "oracolo di Kevin Bacon". Partito come un gioco da parte di alcuni giovani studenti americani, si è dimostrato un efficace metodo di analisi di una rete sociale. L'oracolo è disponibile come risorsa internet (www.cs.virginia.edu/oracle) e permette di calcolare i gradi di separazione tra un qualsiasi attore e Kevin Bacon. I gradi di separazione sono espressi in termini di partecipazioni comuni a film con Kevin Bacon o partecipazioni con attori che hanno girato film con questo attore. Il sorprendente risultato è che la maggior parte degli attori è collegato con pochi link a Bacon. Ma questo fatto è dovuto ad una particolare prolificità produttiva di Kevin Bacon o è generalizzabile? Nella versione aggiornata dell'oracolo è ora possibile cercare le connessioni tra due attori qualunque. Facendo diverse prove si ottiene lo stesso risultato ottenuto con Bacon: due attori qualunque sono connessi da pochi gradi di separazione, in media 2,79.

Un esempio simile è dato nel mondo scientifico, in particolare dall'ambito matematico, dal cosiddetto "numero di Erdős". Paul Erdős è un geniale matematico ungherese, che ha condotto una vita assolutamente non convenzionale (Hoffman, 1999) fuori dagli schemi universitari canonici, durante la quale ha collaborato con un elevatissimo numero di colleghi. Data la sua genialità, molti dei suoi interventi portarono a pubblicazioni, e l'eccentricità di Erdős ha fatto sì che queste fossero distribuite nella comunità scientifica. Il numero di Erdős rappresenta una misura dei gradi di separazione nella comunità scientifica. Avere un numero di Erdős pari a 1 significa aver collaborato direttamente con lui alla pubblicazione di un lavoro; un numero di Erdős pari a 2 significa aver collaborato con un suo coautore, e così via. Dall'analisi dei numeri di Erdős è emerso che la maggior parte dei matematici ha un basso numero di connessioni, che oscilla tra 2 e 5 (De Castro e Grossman, 1999): anche la rete delle collaborazioni tra matematici è un piccolo-mondo.

Oltre che nelle reti sociali, questa proprietà è stata riscontrata anche in altre reti: ad esempio il Web, una rete i cui nodi sono costituiti dalle pagine, connesse da milioni

di link. Albert, Jeong e Barabási (1999) hanno studiato il Web con l'obiettivo di scoprire la sua topologia e calcolare i gradi di separazione tra le pagine. Con questa ricerca hanno calcolato una distanza media tra le pagine di 11 passaggi. Il collegamento tra le pagine web è costituito dagli URL (localizzatori di risorse uniformi): ogni pagina web contiene link ad altre pagine. Con un *click* su un URL possiamo muoverci da una pagina all'altra. La presenza di 11 gradi di separazione significa che per passare da una qualunque pagina web ad un'altra scelta a caso sono necessari in media 11 *click*. Studi successivi hanno rivelato un basso grado di separazione in molte altre reti, come le reti alimentari, la rete neurale di *Caenorhabditis elegans* (Watts e Strogatz, 1998), e le reti di interazione tra le molecole all'interno della cellula (Jeong et al., 2001; Wagner e Fell, 2001). La media dei gradi di separazione in queste reti varia tra 2 e 14.

Il motivo dell'esistenza della proprietà di piccolo mondo nelle reti reali è legato alla natura altamente interconnessa delle reti stesse. Nel modello casuale, si è visto, è sufficiente un solo link per nodo perché si formi un gruppo di nodi gigante. La connettività media delle reti reali supera questo valore minimo: l'aggiunta di altri link provoca il collasso della distanza tra i nodi e perciò la presenza di un basso grado di separazione. (Barabási, 2004).

Il basso numero di gradi di separazione tra i nodi è una caratteristica dei grafi casuali riscontrata in molte reti reali. In realtà, però, il modello dei grafi casuali non è in grado di spiegare il funzionamento delle reti reali. Infatti dall'analisi di diverse reti, come il World Wide Web e la stessa rete sociale, sono emerse proprietà non condivise dai grafi casuali:

- il clustering
- la distribuzione dei gradi dei nodi secondo una legge di potenza (n.b. per grado di un nodo si intende il numero di link posseduti da quel nodo).

1.4. Il clustering

Un'importante proprietà della rete sociale è il fatto che essa è suddivisa in un gran numero di gruppi di nodi altamente connessi tra loro. Le persone infatti non sono collegate a caso, ma nelle reti sociali contano anche i nuclei di relazione locali, formati da persone con interessi simili, o più vicine geograficamente. In altre parole, in una rete sociale gli amici di una persona hanno una maggiore probabilità di essere

amici a loro volta, rispetto a due persone qualunque. Questa proprietà è chiamata *clustering*, e non viene osservata nei grafi casuali: qui, infatti, la presenza di un legame è indipendente dalla presenza di altri legami. Si può quindi immaginare la rete come formata da una serie di sottografi completi, all'interno dei quali tutti i nodi sono connessi, detti *cluster*. Questi cluster sono collegati tra loro da pochi legami tra nodi di gruppi diversi. Si possono quindi definire due tipi di legami all'interno di una rete: i legami forti, che si instaurano tra i nodi all'interno di un cluster, e i legami deboli, che connettono nodi di cluster diversi, e sono responsabili della connettività dell'intera rete. Rimuovere un legame debole, infatti, significa isolare i cluster ed aumentare la distanza tra i nodi. I nodi A e B nella figura 2 sono connessi da un legame debole. Se eliminassimo questo legame, la distanza tra i due nodi aumenterebbe: per passare da A a B dovremmo percorrere tutto il grafo.

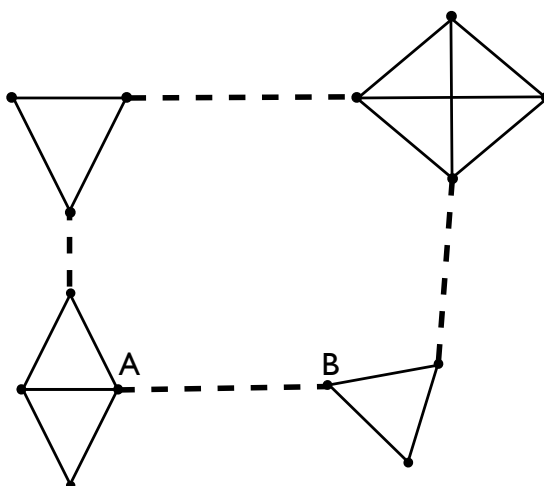


Fig. 2. Rappresentazione di una rete formata da 4 cluster connessi da altrettanti "ponti sociali" formati dai legami deboli (linea tratteggiata).

Questo tipo di rappresentazione di una rete e quindi della società è stato proposto nel 1973 da Mark Granovetter, un sociologo americano, nel suo articolo "The strength of weak ties" (La forza dei legami deboli) (Granovetter, 1973; 1983). In questo lavoro Granovetter sostiene che in una rete sociale i legami più importanti sono quelli deboli, vale a dire i legami che costruiscono ponti tra gruppi diversi. Dagli studi di Granovetter è emerso infatti che in molte situazioni sociali, come la ricerca di un nuovo lavoro, i legami deboli contano più di quelli forti. All'interno della propria ristretta cerchia di amici e parenti, infatti, le informazioni si diffondono meno,

perché tutti si conoscono. Se invece le informazioni vengono diffuse attraverso un legame debole, hanno un'alta probabilità di raggiungere più persone. Granovetter intervistò diverse persone assunte di recente e analizzò il loro rapporto con la persona che le aveva aiutate a trovare lavoro. Solo il 16% delle persone aveva avuto il lavoro attraverso un legame forte; nell'84% dei casi, invece, erano state persone che incontravano occasionalmente o raramente, con cui cioè avevano legami deboli, a fornire loro il contatto lavorativo.

I grafi casuali di cui abbiamo parlato in precedenza non presentano il clustering, e non possono quindi essere usati per modellizzare le reti reali, pur condividendo con esse la proprietà di piccolo mondo. I grafi regolari, a loro volta, possiedono nuclei di relazioni come quelli che si rinvengono nelle reti sociali, ma non la proprietà di piccolo mondo (Buchanan, 2003). È necessario quindi un diverso modello matematico che comprenda entrambe le proprietà, il clustering e il piccolo mondo. Nel 1998 Watts e Strogatz hanno proposto un nuovo modello, quello delle reti piccolo-mondo, che possiede caratteristiche intermedie tra un grafo regolare e un grafo casuale. Il modello verrà descritto nella seconda parte della trattazione.

1.5. La distribuzione dei gradi dei nodi secondo una legge di potenza

Dopo la pubblicazione da parte di Watts e Strogatz del modello delle reti piccolo-mondo, dall'analisi di altre reti reali, come il web (Albert et al., 1999), la rete di interazioni tra le proteine (Vogelstein et al., 2000), la rete metabolica (Jeong et al., 2000), sono emerse nuove caratteristiche, che non rientrano in nessuno dei modelli di rete precedenti.

Innanzitutto, in queste reti i gradi dei nodi (cioè il numero dei link per ogni nodo) si distribuiscono secondo una legge della potenza, e non secondo una distribuzione poissoniana, come previsto dal modello di rete casuale. Questo fa sì che ci siano molti nodi poco connessi, e pochi nodi con molti link. Questi nodi iperconnessi vengono definiti "hub". Gli hub hanno un ruolo cruciale, in quanto determinano la stabilità strutturale, il comportamento dinamico, la robustezza e la tolleranza ad attacchi ed errori della rete. È stato quindi proposto un terzo modello di rete, quello delle reti a invarianza di scala.

1.6. Nuovi modelli di rete

La scoperta del clustering e della presenza degli hub nelle reti reali ha condotto quindi all'abbandono del modello delle reti casuali a favore di modelli più realistici.

In particolare, i modelli di rete che vengono utilizzati attualmente sono due:

- Il modello piccolo-mondo, proposto da Watts e Strogatz nel 1998, che descrive una rete a metà strada tra un grafo casuale e un grafo regolare, e tiene conto del clustering;
- Il modello delle reti a invarianza di scala, di Barabási e Albert (1999), che spiega le reti regolate dalla legge di potenza.

Nella prossima sezione verranno descritti i tre modelli di rete presentati, casuale, piccolo-mondo e a invarianza di scala, nei loro aspetti matematici.

2. MODELLI DI RETE

2.1. Il modello dei grafi casuali

2.1.1. Introduzione

La teoria dei grafi casuali fu introdotta da Paul Erdős e Alfréd Rényi negli anni '50 del secolo scorso. I due matematici definiscono un grafo casuale come N nodi connessi da n archi, che sono scelti casualmente tra gli $N(N-1)/2$ archi possibili (Erdős e Rényi, 1959; Albert e Barabási, 2002). In totale ci saranno $\binom{n}{N(N-1)/2}$ grafi con N nodi e n archi.

È possibile definire i grafi casuali anche con un modello binomiale. Si parte da N nodi, di cui ogni coppia è connessa con probabilità p . Il numero totale di archi è una variabile casuale con valore atteso $E=p[N(N-1)/2]$. La probabilità di ottenere un grafo G_0 con nodi P_1, P_2, \dots, P_N e n archi sarà $P(G_0)=p^n(1-p)^{N(N-1)/2-n}$ (Albert e Barabási, 2002). Essendo una distribuzione binomiale, la presenza/assenza di un arco è indipendente da quella degli altri archi. Dato che il numero medio di archi nel grafo è pari a $[N(N-1)/2]p$, e che ogni arco connette 2 nodi, il grado medio di un vertice, $\langle k \rangle$, sarà $N(N-1)p/N=(N-1)p$, che per N grande si approssima a Np (Pržulj, 2004).

Erdős e Rényi studiarono diverse proprietà dei grafi casuali, notando che queste non emergono gradualmente, ma in modo improvviso. Si può affermare che un grafo casuale possiede una certa proprietà Q se la probabilità di avere questa proprietà, $P(Q)$, soddisfa il limite

$$\lim_{n \rightarrow \infty} P(Q) = 1$$

Studiarono quindi la p critica a cui il grafo acquisisce una certa proprietà; in particolare, la probabilità che un grafo con N nodi e probabilità di connessione $p=p(N)$ abbia una proprietà Q soddisfa il seguente limite:

$$\lim_{N \rightarrow \infty} P_{N,p(Q)} = \begin{cases} 0 & \text{se } \frac{p(N)}{pc(N)} \rightarrow 0, \text{ cioè } p(N) < pc(N) \\ 1 & \text{se } \frac{p(N)}{pc(N)} \rightarrow \infty, \text{ cioè } p(N) > pc(N) \end{cases}$$

Una proprietà studiata dai due matematici è il variare della topologia del grafo casuale in funzione del numero di archi, quindi al variare di n . Si è visto che per pochi archi (quindi n piccolo) il grafo è frammentato in molte piccole componenti, costituite da un numero di vertici al massimo $O(\log N)$. All'aumentare di n le componenti crescono, collegandosi a nodi isolati e fondendosi ad altre componenti. Per $n=N/2$ avviene una transizione, per cui molti clusters si collegano spontaneamente per formare un'unica componente, chiamata cluster gigante (n.b. per chiarezza si ricorda che questo termine non si riferisce alla proprietà di clustering citata in precedenza e di cui si parlerà in seguito), che è formata da un set di vertici di dimensione molto maggiore rispetto a quello delle altre componenti (Pržulj, 2004). Si parla di transizione di fase, in quanto prima si ha un gruppo di piccoli cluster isolati, poi, aggiungendo un numero di link tale per cui ogni nodo ne abbia almeno uno, si ottiene un cluster gigante in cui, praticamente, tutti i nodi sono connessi. Ciò significa che è sufficiente un solo link per nodo (cioè $\langle k \rangle$, il grado medio di un nodo, si approssima a $Np=1$), perché tutta la rete sia connessa. Si deduce quindi una rilevante proprietà dei grafi casuali: sono delle reti altamente connesse (Barabási, 2004).

2.1.2. La distribuzione dei gradi dei nodi

Dato che in un grafo casuale i link vengono disposti a caso, ogni nodo avrà la stessa probabilità di acquisire un link. Questo porterebbe ad aspettarsi che alcuni nodi avranno più link rispetto ad altri. Erdős e Rényi hanno però dimostrato che se la rete è abbastanza grande, i nodi avranno all'incirca lo stesso numero di link. In un grafo casuale con probabilità di connessione p , il grado k_i di un nodo i , segue una distribuzione binomiale con parametri $N-1$ e p :

$$P(k_i = k) = C_{N-1}^k p^k (1-p)^{N-1-k}$$

Questa probabilità rappresenta il numero di modi in cui k archi possono essere disegnati a partire da un certo nodo.

Si può dimostrare (Albert e Barabási, 2002) che la degree distribution di un grafo casuale per N grandi si approssima ad una distribuzione di Poisson:

$$P(k) \cong e^{-pN} \frac{(pN)^k}{k!} = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

Questa distribuzione presenta un punto massimo, che indica che tutti i nodi hanno perlopiù un numero medio di link. All'aumentare e al diminuire di k la distribuzione decresce rapidamente, dimostrando la rarità delle deviazioni dalla media del numero di link per nodo.

2.1.3. Diametro

Il diametro di un grafo è la distanza massima (espressa in numero di link) tra ogni coppia dei suoi nodi (Albert e Barabási, 2002). I grafi casuali tendono ad avere un basso diametro, anche se il numero dei nodi è molto elevato. Infatti si è visto che quando $\langle k \rangle$ supera la soglia di un link per nodo, la distanza collassa. Data una rete con un grado medio dei nodi pari a k , il numero dei nodi a distanza d da un dato nodo sarà circa $\langle k \rangle^d$. Albert e Barabási (2002) hanno ricavato la seguente formula per calcolare il diametro di un grafo:

$$d = \frac{\ln(N)}{\ln(\langle k \rangle)}$$

Quindi il diametro di un grafo dipende dal logaritmo del numero dei nodi: all'aumentare di N , il diametro crescerebbe di poco; è questo il motivo del basso diametro delle reti molto grandi.

2.2. Il modello delle reti piccolo-mondo

2.2.1. Introduzione

Nel 1998 Watts e Strogatz proposero un modello alternativo, a metà strada tra il modello casuale e la rete regolare, che consentiva di spiegare la tendenza delle reti reali al clustering e contemporaneamente il loro basso grado di separazione.

Watts e Strogatz partirono da una rete circolare ordinata, in cui ogni nodo aveva un legame con i quattro nodi più vicini. In seguito introdussero alcune connessioni tra nodi presi a caso, con probabilità p . Al variare di p è possibile monitorare il passag-

gio da una rete regolare ($p=0$) a una rete casuale ($p=1$) (Albert e Barabási, 2002). Per valori intermedi di p , quello che si ottiene è un grafo piccolo-mondo (Figura 3).

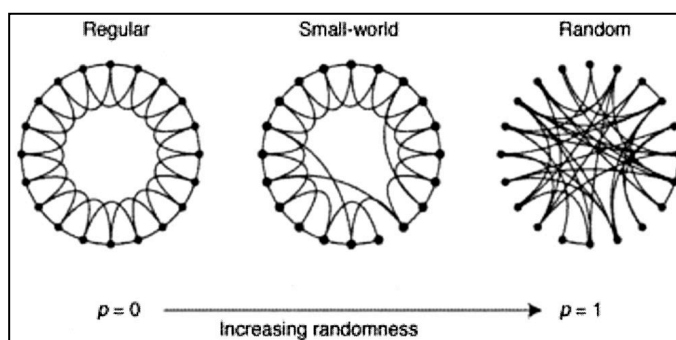


Fig. 3. La procedura utilizzata da Watts e Strogatz per passare da un grafo regolare a uno casuale, ottenendo un grafo piccolo-mondo (da Watts e Strogatz, 1998).

2.2.2. Cammino medio tra i nodi

È possibile studiare come varia il cammino medio tra le coppie di nodi nella rete, $L(p)$, al variare di p (vd. figura 4). Per p piccole, $L(p)$ varia linearmente con la dimensione del sistema, mentre per p grandi la variazione è logaritmica. La ragione di questa brusca caduta di $L(p)$ è da attribuire alla presenza dei link a lungo raggio aggiunti casualmente alla rete regolare. È sufficiente l'aggiunta di un numero relativamente piccolo di queste scorciatoie per ridurre drasticamente il cammino medio tra i nodi, anche se localmente la rete rimane altamente ordinata (Watts e Strogatz, 1998).

2.2.3. Il coefficiente di clustering

Si è visto che nelle reti reali, a differenza di quelle casuali, c'è la tendenza a formare dei cluster, piccoli gruppi di nodi altamente connessi tra loro. Si dice che una rete possiede delle caratteristiche di clustering se la probabilità che due nodi siano connessi è maggiore se essi sono connessi ad un terzo nodo. Il clustering è stato descritto per la prima volta da Watts e Strogatz, che hanno ideato un coefficiente in grado di misurare questa tendenza dei nodi della rete a raggrupparsi. Dal punto di vista matematico, il coefficiente di clustering rappresenta il rapporto tra il numero di link esistenti effettivamente tra i k_i nodi adiacenti al nodo i (E_i) e il numero totale di link possibili tra essi ($k_i(k_i-1)/2$), cioè per un nodo i ,

$$C_i = \frac{2E_i}{k_i(k_i - 1)}$$

mentre per tutta la rete il coefficiente di clustering sarà la media di tutti i C_i calcolati per ogni nodo. Il coefficiente di clustering può assumere valori da 0 a 1: un valore vicino a 1 sta ad indicare che i nodi prossimi al nodo i sono connessi anche tra loro, mentre un valore vicino allo 0 indica che i vicini al nodo i non sono connessi tra loro, cioè il nodo i è l'unico che li tiene legati.

In una rete casuale il coefficiente di clustering è $C_{rand}=p=\langle k \rangle / N$, cioè la probabilità che due vicini di un nodo siano connessi tra loro è uguale alla probabilità che due nodi scelti a caso siano connessi tra loro.

In una rete piccolo-mondo il valore del coefficiente di clustering, per piccoli valori di p è simile a quello delle reti regolari, cioè

$$C = \frac{3(k - 2d)}{4(k - d)}$$

Con k =numero dei nodi adiacenti a un nodo, d =dimensione del grafo, che per k grandi tende a $3/4$ (Newman, 2000). Per valori di p tendenti a 1, il coefficiente di clustering delle reti piccolo-mondo si approssima a $C_{rand}=k/n \ll 1$ (vd. figura 4).

In conclusione, Watts e Strogatz hanno rilevato l'esistenza di un ampio intervallo di p in cui $L(p)$ è piccolo come L_{rand} (basso grado di separazione tra i nodi) e $C(p) \gg C_{rand}$ (alto grado di clusterizzazione). Questo dimostra l'esistenza di una rete i cui nodi sono altamente raggruppati e contemporaneamente vi è un breve cammino medio tra i nodi: la rete piccolo-mondo.

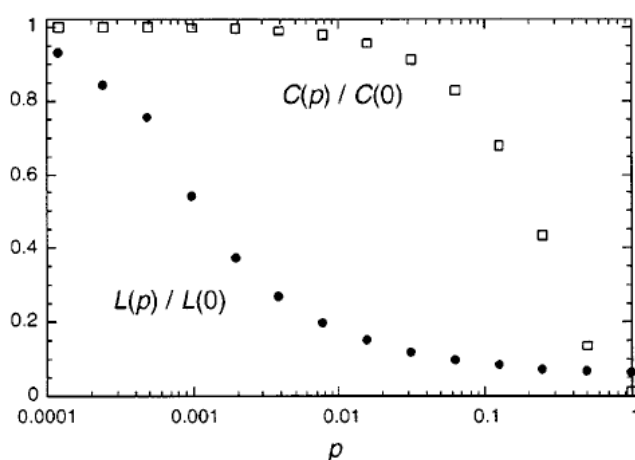


Fig. 4. Andamento del cammino medio tra i nodi, L , e del coefficiente di clustering, C , in funzione di p , in una rete piccolo-mondo (da Watts e Strogatz, 1998).

Queste proprietà sono state ritrovate in diverse reti reali, come la rete degli attori, la rete elettrica, la rete neurale di *Caenorhabditis elegans*, il World Wide Web, la rete metabolica di *Escherichia coli* e molte altre, come mostrato in tabella (Albert e Barabási, 2002).

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook <i>et al.</i> , 2001a, Pastor-Satorras <i>et al.</i> , 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási <i>et al.</i> , 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási <i>et al.</i> , 2001	9
<i>E. coli</i> , substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
<i>E. coli</i> , reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook <i>et al.</i> , 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
<i>C. Elegans</i>	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

In tutte queste reti si può vedere che il cammino medio tra i nodi è comparabile con quello di una rete casuale con la stessa dimensione e grado dei nodi, mentre il coefficiente di clustering è maggiore di quello della rete casuale.

2.3. Il modello delle reti a invarianza di scala

2.3.1. Introduzione

Analizzando diverse reti reali, quali il World Wide Web, la rete delle citazioni scientifiche, la rete degli attori, Barabási e Albert (1999) hanno evidenziato una caratteristica importante: in queste reti i gradi dei nodi si distribuiscono in modo diverso rispetto a quanto previsto dai modelli dei grafi casuali (di Erdős e Rényi, 1959) e piccolo-mondo (di Watts e Strogatz, 1998). In queste reti, infatti, la probabilità che un nodo interagisca con k altri nodi decade secondo una legge di potenza, cioè $P(k) \sim k^{-\gamma}$. L'esponente di grado della distribuzione, γ , assume valori diversi a seconda della rete, compresi in un intervallo tra 2.1 e 4. Ciò significa che ogni volta che il numero di connessioni raddoppia, il numero di nodi con quel numero di connessioni diminuisce di 2^γ volte. Questo spiega l'esistenza, in queste reti, di pochi nodi altamente connessi, gli hub, e molti nodi con pochi link.

Una distribuzione secondo la legge di potenza non è rappresentata da una curva con un picco, come può essere una distribuzione a campana, ma da una curva decrescente con continuità (Figura 5). Nel caso delle reti, questo significa che molti nodi con poche connessioni coesistono con pochi nodi altamente connessi. Ed è questa la principale differenza tra le reti reali e le reti casuali. In queste ultime, infatti, la distribuzione dei gradi dei nodi segue una distribuzione normale: esiste un picco, che mostra come la maggior parte dei nodi abbia lo stesso numero di link, mentre le deviazioni dalla media, in entrambi i sensi, sono rare.

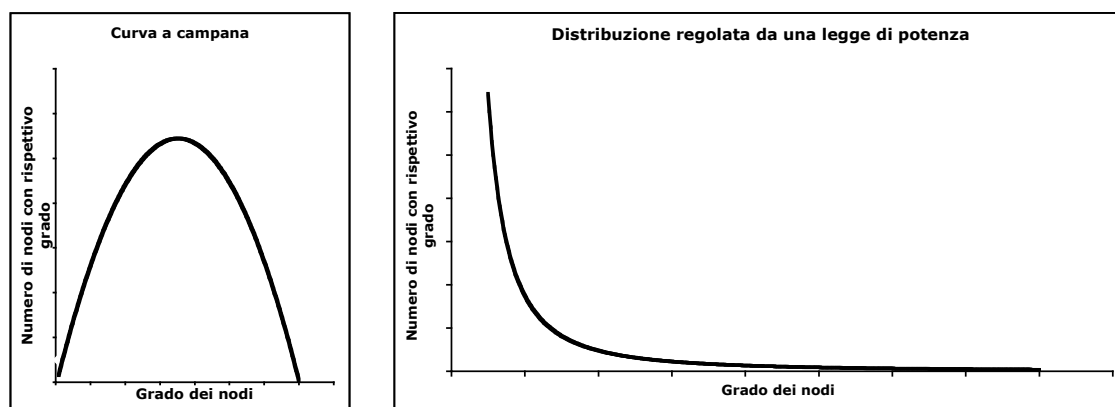


Fig. 5. Esempi di distribuzione a campana e distribuzione secondo una legge di potenza.

In una serie di dati che si distribuisce a campana troveremo sempre valori vicini alla media, mentre valori che si discostano da essa sono inattesi. Una rete casuale, quindi, ha una scala, rappresentata dal nodo medio, il nodo caratteristico. Nelle reti con distribuzione secondo la legge di potenza, invece, l'assenza di un picco comporta l'assenza di un nodo medio: è presente invece una gerarchia continua di nodi, dai rari hub ai numerosi nodi poco connessi (Barabási, 2004). Da questa rappresentazione emerge una importante caratteristica: non esiste un valore medio globale, ma si osserva solo che più un nodo è connesso, più è raro.

Le reti a invarianza di scala hanno la caratteristica dell'autosimilarità: ogni parte della rete è simile alla rete complessiva. Queste reti, quindi, non hanno una scala intrinseca, non possiamo isolare un nodo rappresentativo di tutti gli altri: per questo sono chiamate "reti a invarianza di scala" (Barabási e Albert, 1999).

Questa caratteristica delle reti reali non può però essere spiegata con i modelli casuale e piccolo-mondo. Questi, infatti, posano su due assunti che sono incompatibili con il modello a invarianza di scala:

- il numero dei nodi è fisso e rimane invariato per la durata di vita della rete: la rete quindi non cresce, non vengono aggiunti nuovi nodi. Il modello delle reti a invarianza di scala, invece, descrive una rete dinamica, che si evolve con l'aggiunta di nuovi nodi ad ogni unità di tempo;
- i nodi sono connessi in modo casuale: la rete che ne risulta è una rete egualitaria, in cui tutti i nodi si equivalgono. La rete a invarianza di scala, invece, è di tipo gerarchico, con nodi poco connessi che convivono con hub ricchi di link.

Per spiegare l'insorgere delle reti a invarianza di scala, infatti, è necessario studiare la nascita e l'evoluzione delle reti, secondo un modello dinamico. Barabási e Albert (1999) propongono due meccanismi che a loro parere sono in grado di spiegare l'esistenza delle reti a invarianza di scala e l'insorgenza della legge di potenza:

- la crescita, quindi il numero dei nodi non è fisso e invariato, ma la rete cresce con l'aggiunta di un nuovo nodo ad ogni intervallo di tempo;
- il collegamento preferenziale: ogni nuovo nodo non si collega in modo casuale a quelli già presenti, ma tende a favorire i nodi con un numero maggiore di connessioni; la probabilità Π che un nuovo nodo, j , scelga di connettersi al nodo i , è proporzionale al grado k_i di questo:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

Queste leggi sono entrambe necessarie perché si formi una rete a invarianza di scala. Infatti, senza il collegamento preferenziale una rete in crescita mostra una distribuzione simile ad una curva a campana, escludendo gli hub; mentre senza crescita, si torna ad un modello statico, che non può generare una legge di potenza.

La rete si accresce quindi secondo il famoso meccanismo dei “ricchi che diventano sempre più ricchi”, responsabile della nascita degli hub. I nodi più vecchi avranno così un numero maggiore di link, dal momento che hanno avuto più tempo per accumularli. Molti esempi di reti reali, però, mostrano che non è sempre così: ci sono casi in cui il grado di un nodo non dipende solo dalla sua età. Il motore di ricerca Google, ad esempio, è stato lanciato nel 1997 e, nonostante fosse l'ultimo arrivato (i.e. erano già presenti e diffusi altri motori di ricerca), è diventato in poco tempo il nodo più grande della rete rappresentata dal World Wide Web. Per spiegare questo fenomeno, Bianconi e Barabási (2001) hanno introdotto il “modello a fitness”, che contempla la competizione tra i nodi nei sistemi complessi. Ogni nodo, cioè, avreb-

be una capacità intrinseca a competere per i link, alle spese degli altri nodi. In questo modello, quindi, ad ogni nodo è attribuito un valore di fitness, η_i , che non cambia nel tempo. Quindi ad ogni intervallo di tempo viene aggiunto un nuovo nodo j alla rete, con fitness η_j . Ogni nuovo nodo si connette con m archi ai nodi già presenti nella rete, e la probabilità di connettersi ad un nodo i è proporzionale al grado e alla fitness del nodo i :

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}.$$

Questa generalizzazione del meccanismo del collegamento preferenziale permette anche ad un nodo relativamente giovane di acquisire un elevato numero di collegamenti, se ha un valore di fitness elevato.

2.3.2. Cammino medio e clustering

Le reti a invarianza di scala presentano un cammino medio tra i nodi più breve rispetto a quello di una rete casuale comparabile, e che si accresce all'aumentare della dimensione della rete. Per quanto riguarda il clustering, hanno un coefficiente di clustering cinque volte maggiore rispetto a quello dei grafi casuali, e questo fattore cresce col numero dei nodi. Questa dipendenza del coefficiente di clustering dalla dimensione della rete differenzia le reti a invarianza di scala da quelle piccolo-mondo.

2.3.3. La tolleranza ad attacchi ed errori

Si è detto che gli hub, i nodi altamente connessi delle reti a invarianza di scala, sono elementi importanti per la topologia e il funzionamento delle reti stesse. In particolare, hanno un ruolo fondamentale nella tolleranza e resistenza della rete nei confronti degli errori o di attacchi esterni. Per errore si intende la rimozione di nodi casualmente, mentre un attacco consiste nella selezione e rimozione di pochi nodi che giocano un ruolo vitale nel mantenere la connettività della rete. Albert, Jeong e Barabási (2000) hanno studiato questo aspetto delle reti a invarianza di scala, confrontandole con le reti a distribuzione esponenziale, spiegate dal modello casuale. La tolleranza agli errori è stata misurata come cambiamento del diametro della rete in seguito alla rimozione casuale di un certo numero di nodi. Il risultato è stato che, mentre nelle reti esponenziali il diametro cresce al crescere del numero dei nodi rimossi, nelle reti a invarianza di scala il diametro non aumenta fino ad un certo level-

lo di errore. Anche con la rimozione del 5% dei nodi, la comunicazione tra gli altri nodi non viene influenzata. Questa tolleranza agli errori è legata alla connettività disomogenea della rete: dato che la maggior parte dei nodi ha pochi link, c'è un'alta probabilità che l'errore colpisca i nodi poco connessi, e quindi che non si abbiano grosse conseguenze sulla connettività totale della rete.

Conoscendo la topologia della rete, però, è possibile programmare degli attacchi mirati, che vadano a colpire direttamente i nodi più connessi. Analizzando le risposte dei due tipi di rete agli attacchi mirati, Albert, Jeong e Barabási hanno in effetti rilevato un rapido aumento del diametro delle reti a invarianza di scala dopo la rimozione dei nodi altamente connessi. I pochi hub sono responsabili dell'alta connettività della rete. Quindi l'alta tolleranza agli errori garantita da questa topologia porta anche ad una vulnerabilità agli attacchi.

Questa proprietà delle reti a invarianza di scala è stata studiata per esempio nella rete di interazione tra le proteine nel lievito *Saccharomyces cerevisiae* (Jeong et al., 2001). Nello stesso modo funzionano anche le reti alimentari: si è visto infatti che un ecosistema è in grado di sopravvivere all'estinzione casuale di alcune specie, mentre l'eliminazione delle specie guida, maggiormente connesse, determina il collasso del sistema. Un caso molto studiato è quello della protezione delle lontre marine in California, in seguito alla loro diminuzione a causa della caccia. L'aumento delle lontre ha determinato la riduzione della popolazione di ricci, di cui si nutrivano; questo ha portato all'aumento delle alghe e, quindi, dei pesci. La protezione di una sola specie ha quindi alterato profondamente l'ecologia, e anche l'economia della zona, che prima era dominata dai molluschi, mentre ora dai pesci (Levin, 1999).

Nonostante questa vulnerabilità agli attacchi delle reti a invarianza di scala, però, per farle collassare del tutto si dovrebbero eliminare contemporaneamente i più grandi hub, evento abbastanza improbabile. In realtà, bisogna tenere conto però anche dei guasti a cascata che, se colpissero i nodi più connessi, danneggerebbero la rete in modo grave.

3. SINTESI

Nelle sezioni precedenti è stata presentata la teoria delle reti e sono stati descritti i modelli matematici che vengono oggi utilizzati per la rappresentazione e lo studio

dei sistemi complessi. In questa sezione verranno riassunti i punti chiave della teoria, utili a comprendere le analisi sulle reti metaboliche esposte in questo lavoro.

A partire dalla metà del secolo scorso, il modello di rete prevalente era quello dei *grafi casuali* di Erdős e Rényi: in questa rete i nodi sono connessi a caso, e la probabilità di avere un link è uguale per tutti i nodi. La distribuzione dei gradi dei nodi viene rappresentata da una curva a campana, con un picco che identifica un nodo medio, rappresentativo di tutta la rete. La rete casuale è caratterizzata da un basso diametro, quindi ogni coppia di nodi è collegata da un numero minimo di passaggi. Questa proprietà, detta di “*piccolo mondo*”, è una caratteristica riscontrata nella maggior parte delle reti reali. In particolare è stata studiata nelle reti sociali, dove è nota con l’espressione “sei gradi di separazione”, che indica che due persone qualsiasi nel mondo possono essere collegate in media da 6 passaggi di conoscenti (i.e. link).

Dall’analisi delle reti reali sono però emerse altre proprietà, non condivise dai grafi casuali spiegati dal modello di Erdős e Rényi. Innanzitutto le reti reali tendono ad essere formate da sottografi, piccoli gruppi di nodi molto connessi. Questi gruppi, o cluster, sono connessi tra loro da pochi legami deboli. I legami deboli sono molto importanti perché mantengono la rete connessa e sono responsabili del suo basso diametro: se fossero eliminati, le distanze tra i nodi appartenenti a cluster diversi aumenterebbero di molto. Dato che le reti casuali non contemplano il clustering, non sono in grado di descrivere le reti reali. Per questo è stato proposto un nuovo modello, quello delle *reti piccolo-mondo* (Watts e Strogatz, 1998), che descrive una rete intermedia tra un grafo casuale (che ha un basso cammino medio tra i nodi ma non il clustering) e un grafo ordinato (che ha il clustering ma non è un piccolo mondo). La rete piccolo-mondo di Watts e Strogatz, quindi, da un lato possiede un basso diametro, dall’altro è formata da una serie di cluster molto connessi legati tra loro da legami deboli: in questo modo descrive meglio le reti reali rispetto al modello casuale. Entrambi questi modelli, però, descrivono delle reti egualitarie, dove tutti i nodi sono uguali, in quanto la loro connessione è casuale.

La scoperta che alcune reti reali hanno nodi altamente connessi, gli *hub*, ha portato alla necessità di un nuovo modello di rete. Il modello delle *reti a invarianza di scala* (Barabasi e Albert, 1999) descrive delle reti in cui la distribuzione dei gradi dei nodi segue una legge di potenza: vi sono quindi molti nodi con pochi link che coesistono con pochi hub ricchi di collegamenti. Questa rete non è più una rete egualitaria, ma

gerarchica. Per spiegare l'insorgenza degli hub, sono stati proposti due meccanismi, la crescita e il collegamento preferenziale, per cui nuovi nodi si collegano con maggiore probabilità ai nodi che hanno più link tra quelli già presenti nella rete.

Nelle reti reali sono quindi state riconosciute proprietà con una precisa organizzazione. Queste proprietà sono comuni a sistemi che in apparenza sono molto diversi tra loro, dalla cellula alla rete sociale, e ciò permette l'applicazione degli stessi principi nello studio e comprensione di queste realtà.

2. Le reti cellulari

Nell'ultimo secolo le ricerche in biologia si sono concentrate sullo studio delle componenti cellulari e delle loro funzioni, in un'ottica riduzionista. Oggi, però, ci si è resi conto che molte caratteristiche della cellula non sono riconducibili ad una sola molecola, ma anzi sono il risultato di interazioni complesse tra le componenti cellulari, quali proteine, DNA, RNA e altre molecole. Si sta quindi passando da un'ottica molecolare ad una modulare, per cui la cellula viene suddivisa in moduli, composti da molti tipi di molecole, le cui funzioni nascono dall'interazione di queste componenti, ma che non possono essere facilmente previste dallo studio delle componenti isolate (Hartwell et al., 1999). Difficilmente infatti possiamo comprendere il comportamento di un organismo vivente a partire dalle sue componenti molecolari, ma dobbiamo considerarlo nella sua integrità, tenendo conto di come queste parti interagiscono tra loro. Un modo per raggiungere questo obiettivo è quello di ricostruire le reti di interazione tra le componenti cellulari.

All'interno della cellula possono essere identificate diverse reti:

- la rete metabolica, i cui nodi sono le sostanze chimiche, connesse da reazioni biochimiche;
- la rete di regolazione: i nodi in questo caso sono i geni e le proteine, mentre i link sono costituiti dalle interazioni biochimiche tra queste componenti;
- la rete di interazione tra proteine, che sono collegate se interagiscono fisicamente o funzionalmente tra loro.

Tutte queste reti non sono indipendenti, ma fanno parte di un'unica rete cellulare, comprendente tutte le componenti della cellula connesse da tutte le interazioni fisiologicamente rilevanti, da quelle biochimiche a quelle fisiche. Se, quindi, con il sequenziamento dei genomi degli esseri viventi, è stato fatto un grosso passo avanti nella conoscenza, ora è necessario un ulteriore passo, che conduca ad una "mappa della vita" (Barabási, 2004), che comprenda tutta la rete di interazione tra le componenti cellulari, per una visione integrata del funzionamento delle cellule.

1. LE RETI METABOLICHE

La rete metabolica di un organismo riassume l'insieme delle reazioni biochimiche che permettono la vita cellulare. Questa rete ha come nodi i composti chimici, lega-

ti da link che rappresentano le reazioni. Diversi studi su queste reti hanno fornito informazioni riguardo la loro architettura.

Jeong et al. (2000) hanno analizzato le reti metaboliche di 43 organismi, appartenenti a diversi gruppi. Hanno costruito una rete i cui nodi sono rappresentati dai substrati, mentre i link sono costituiti dalle reazioni metaboliche, fisicamente costituite dai complessi temporanei formati dai reagenti delle reazioni. Hanno quindi analizzato la struttura di questa rete. Dai risultati è emerso che la probabilità che un dato substrato partecipi a k reazioni è distribuita secondo una legge di potenza: la rete metabolica, quindi è una rete a invarianza di scala, e questo è stato dimostrato per tutti i 43 organismi. Essendo una rete ordinata, è stata analizzata la distribuzione dei gradi dei nodi in entrata e in uscita: in entrambi i casi la distribuzione segue una legge di potenza, con esponenti γ_{in} e γ_{out} compresi tra 2.0 e 2.4. Per quanto riguarda la proprietà di piccolo mondo, è stato considerato come diametro della rete il pathway biochimico più breve esistente tra tutte le coppie di substrati. Anche in questo caso il risultato è stato omogeneo: in tutti gli organismi, indipendentemente dal loro grado di complessità, la rete metabolica è un piccolo mondo, con un cammino medio tra i nodi intorno a 3 passaggi. Questo aspetto è molto importante, in quanto se le molecole fossero separate da un numero elevato di passaggi, qualsiasi cambiamento nella concentrazione della prima molecola dovrebbe passare attraverso molte reazioni intermedie prima di raggiungere la seconda molecola, e quindi ogni perturbazione finirebbe per decadere. Invece in una rete piccolo mondo in cui le molecole sono separate da un massimo di 3 passaggi, le perturbazioni non restano localizzate, ma ogni cambiamento nella concentrazione di una molecola raggiungerebbe rapidamente tutte le altre. Infine Jeong e collaboratori hanno studiato la rete dal punto di vista della tolleranza agli attacchi ed errori: come dimostrato dal modello di rete a invarianza di scala (Albert et al., 2000), la rete metabolica si è mostrata tollerante nei confronti degli errori casuali, ma vulnerabile agli attacchi mirati verso gli hub, i substrati che partecipano a molte reazioni. Questi hub, in genere, sono anche i soli substrati che sono presenti in tutti gli organismi analizzati. In conclusione questo lavoro ha mostrato che la rete metabolica è una rete a invarianza di scala e che presenta proprietà di piccolo mondo; inoltre, queste caratteristiche sono conservate dal punto di vista evolutivo in quanto condivise da archebatteri, eubatteri ed eucarioti.

Risultati simili sono stati ottenuti da Wagner e Fell (2001). In particolare la loro attenzione si è rivolta ad un unico organismo, *Escherichia coli*. Hanno considerato due tipi di rete, la rete dei substrati, in cui i composti della rete sono collegati se partecipano alla stessa reazione biochimica; e la rete delle reazioni, in cui i nodi della rete (le reazioni metaboliche) sono connessi se condividono almeno un composto. Di entrambe queste reti hanno calcolato i gradi di ogni nodo e la loro distribuzione, il cammino medio tra due nodi e il coefficiente di clustering. Analizzando la distribuzione dei gradi dei nodi hanno evidenziato che la rete dei substrati è del tipo a invarianza di scala, mentre la distribuzione dei gradi dei nodi della rete delle reazioni non segue una legge di potenza, ma quest'ultima si avvicina maggiormente al modello della rete piccolo-mondo di Watts e Strogatz. In particolare hanno studiato due versioni della rete metabolica di *E.coli*: una completa, e una in cui sono stati rimossi i nodi con grado più alto, come ATP, ADP e NAD, perché si suppone siano collegati a quasi tutti gli altri metaboliti. In entrambe le reti è stata riscontrata la proprietà di piccolo mondo, con un diametro che varia da 2.9, per la rete completa a 3.9 per la rete priva dei nodi iperconnessi, e un alto coefficiente di clustering.

Wagner e Fell, inoltre, danno una possibile spiegazione per l'origine evolutiva delle proprietà delle reti. In particolare, essi sostengono che, in base al meccanismo del collegamento preferenziale (Albert e Barabási, 1999), i nodi più connessi della rete sarebbero quelli più antichi. Infatti hanno scoperto che le molecole maggiormente connesse nella rete metabolica sono quelle più antiche dal punto di vista evolutivo. Suppongono inoltre che la proprietà di piccolo mondo delle reti sia una conseguenza della selezione naturale, che avrebbe favorito le reti in cui il tempo di transizione tra gli stati metabolici viene minimizzato.

Un terzo lavoro riguardo alle reti metaboliche è quello di Ma e Zeng (2003). Gli autori criticano però la scelta di Jeong et al. (2000) di considerare, come nodi della loro rete metabolica, anche i metaboliti ricorrenti, come ATP, ADP, NADH, NAD⁺, che essendo presenti nella maggior parte delle reazioni, rendono poco significativo il calcolo del diametro della rete. Essi quindi costruiscono la rete metabolica distinguendo i metaboliti ricorrenti da quelli normali, e tenendo conto della reversibilità delle reazioni. Ma e Zeng hanno ricostruito le reti metaboliche di 80 organismi, a partire dai loro genomi sequenziati. Hanno poi analizzato le reti statisticamente. Sia la distribuzione dei gradi dei nodi in entrata che quella in uscita se-

guono una legge della potenza, dimostrando che anche la rete metabolica priva dei metaboliti ricorrenti è una rete a invarianza di scala. Per quanto riguarda gli hub, sono stati identificati 20 metaboliti con un grado elevato nella maggior parte degli organismi: tra questi figurano glicerato-3-fosfato, piruvato, acetil-CoA e altri. Particolarmente importante è l'universalità di questi metaboliti, che sono degli hub per le reti metaboliche di organismi diversi. Calcolando il percorso medio tra due nodi della rete e il diametro (distanza massima tra due nodi), Ma e Zeng hanno rilevato delle differenze tra i diversi organismi, diversamente da quanto misurato da Jeong et al. (2000). Il cammino medio tra due nodi tende a crescere con la dimensione della rete, soprattutto nel caso di reti di piccole dimensioni. Gli autori concludono perciò che nonostante la struttura di base della rete metabolica sia simile per tutti gli organismi, essa mostra però differenze quantitative: è probabile che queste differenze riflettano le diverse storie evolutive degli organismi.

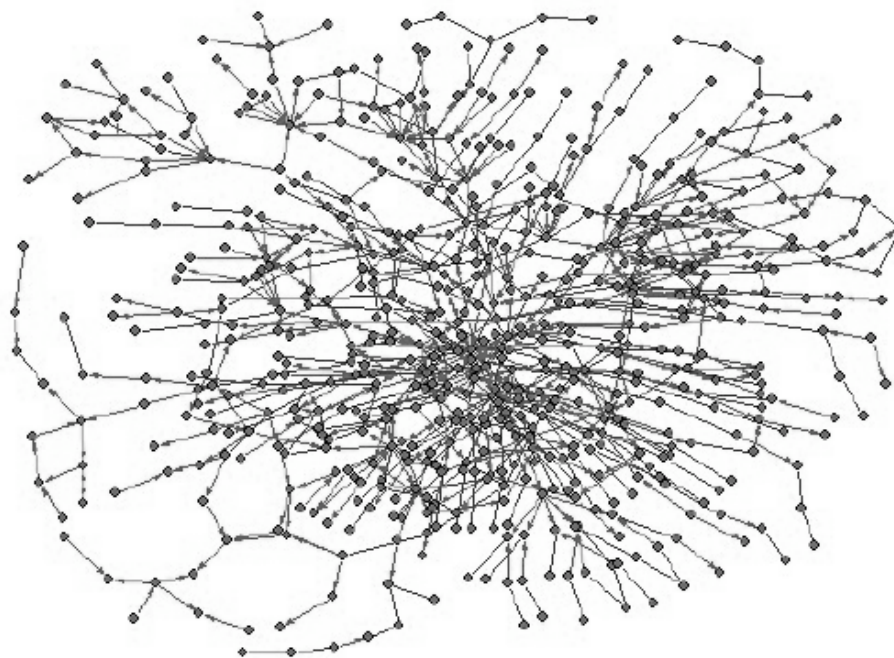


Fig. 6. La struttura della rete metabolica di *E.coli* ricostruita da Ma e Zeng (2003).

Tutti questi studi analizzati (Jeong et al., 2000, Wagner e Fell, 2001, Ma e Zeng, 2003) hanno quindi dimostrato che la rete metabolica è una rete a invarianza di scala, con proprietà di piccolo mondo: a seconda del metodo utilizzato per costruire la rete, sono stati identificati diversi hub, sostanze iperconnesse con un ruolo importante per la connettività della rete stessa, come riassunto in tabella.

Tab. 1. Elenco dei primi 10 hub (composti con grado maggiore) identificati nella rete metabolica di *E.coli* nei diversi studi analizzati.

Jeong et al. (2000)	Wagner e Fell (2001)	Ma e Zeng (2003)
H ₂ O	L-glutamato	Glicerato 3P
ADP	piruvato	D-ribosio ₅ P
P	CoA	Acetil-CoA
ATP	α-chetoglutarato	piruvato
L-glutamato	L-glutammina	D-xilulosio ₅ P
NADP ⁺	L-aspartato	D-fruttosio 6P
PP	Acetil-CoA	5P-D-ribosio 1P
NAD ⁺	fosforibosilPP	L-glutamato
NADPH	tetraidrofolato	D-gliceraldeide 3P
NADH	succinato	L-aspartato

Un altro lavoro sull'argomento, che però contrasta con quelli precedenti, è quello di Arita (2004). In questo studio, l'autore costruisce la rete metabolica di *E.coli* collegando i metaboliti non in base a legami funzionali (partecipazione alla stessa reazione biochimica), ma strutturali. Un pathway metabolico tra due metaboliti A e B, in questo lavoro è definito come la sequenza di reazioni biochimiche attraverso le quali almeno un atomo di carbonio di A passa a B. La rete così concepita è stata quindi rappresentata attraverso un grafo orientato. In seguito è stato calcolato il cammino medio tra i nodi: il valore ottenuto, circa 8, è risultato molto maggiore rispetto ad un grafo casuale comparabile. Perciò l'autore ha concluso che se si considerano i rapporti strutturali tra i metaboliti, la rete metabolica di *E.coli* non è un piccolo mondo. Questo metodo di rappresentazione del metabolismo cellulare, però, non può essere utilizzato per valutare l'invarianza di scala o stimare l'importanza evolutiva dei metaboliti, in quanto tiene conto solo degli aspetti strutturali e non della frequenza di utilizzo di un dato substrato.

2. LE RETI DI INTERAZIONE TRA PROTEINE

Nelle reti di interazione tra proteine due nodi (proteine) sono connessi se esiste un legame fisico tra le proteine. Sono stati sviluppati diversi metodi per l'individuazio-

ne delle interazioni tra proteine, sia in vivo, come il sistema del doppio ibrido, la misura della correlazione nell'espressione dell'mRNA, e delle interazioni geniche, che previsioni in silico dall'analisi dei genomi (Przulj, 2004).

Jeong et al. (2001) hanno analizzato la rete di interazione tra le proteine nel lievito *S.cerevisiae*: si tratta di una rete composta da 1870 nodi, costituiti dalle proteine, collegate da 2240 link, cioè interazioni fisiche. La distribuzione dei gradi dei nodi segue una legge della potenza, quindi anche in questo caso la rete è a invarianza di scala. Analizzando la tolleranza ad attacchi ed errori della rete, si è visto che mutazioni casuali nel genoma del lievito non hanno grandi effetti sulla topologia della rete, perché colpendo a caso, è maggiore la probabilità che colpiscano proteine poco connesse. Inoltre la probabilità che la rimozione di una proteina determini effetti letali è correlata col numero di interazioni che la proteina ha: le proteine più connesse sono quelle più importanti per la sopravvivenza della cellula. Risultati simili sono stati ottenuti anche dallo studio della rete di interazione proteica di *Helicobacter pilori* (Rain et al, 2001), e di *Drosophila melanogaster* (Giot et al., 2003).

La topologia a invarianza di scala è quindi una caratteristica che si ritrova a diversi livelli di organizzazione nella cellula, come mostrato dalle reti metaboliche e da quelle proteiche. È probabile che questa caratteristica della cellula sia il risultato di un processo molto frequente: la duplicazione genica (Solè et al., 2002, Vázquez et al., 2003). Con la duplicazione genica si producono due geni, quindi due proteine identiche; nella rete compare quindi un nuovo nodo, che interagirà con le proteine legate alla proteina madre. Ogni proteina, in questo modo, guadagnerà un legame in più e, di conseguenza, le proteine più connesse godranno di un maggior vantaggio, essendo più probabile che la nuova proteina si colleghi ad esse. Quindi il meccanismo della duplicazione genica da una parte consente la crescita della rete, dall'altra porta al collegamento preferenziale.

Lo studio della cellula dalla prospettiva delle reti è molto importante. Ad esempio può consentire di prevedere l'effetto fenotipico della rimozione o disattivazione di un gene, e quindi di individuare i geni e le proteine target per la terapia farmacologica. All'interno della rete cellulare, inoltre, le proteine sono considerate non più per la loro funzione individuale, ma sono suddivise in classi funzionali basate sul loro ruolo cellulare. In questo modo è possibile, a partire dai dati genomici, caratte-

rizzare anche quei prodotti cui non è ancora stata assegnata una funzione (Jeong et al., 2003).

Un esempio di come la teoria delle reti possa essere utile per lo studio della cellula riguarda la proteina p53. Questa proteina, codificata dal gene p53, è un oncosoppressore: normalmente è disattiva, ma viene attivata quando le cellule sono stressate o danneggiate. La sua attivazione può bloccare la moltiplicazione di queste cellule oppure può provocare apoptosi. Si è visto che nel caso dei tumori p53 è disattivata o mutata. Nonostante la scoperta di questa proteina e del suo importante ruolo, si è presto capito che non è sufficiente studiare la singola proteina per sconfiggere i tumori, ma essa va considerata all'interno della complessa rete di interazione e segnalazione cellulare (Vogelstein et al., 2000). In questa rete ci si è subito resi conto che p53 è un hub, con moltissimi legami e, di conseguenza, è anche molto sensibile agli attacchi: un attacco ad un hub, infatti, può essere letale per la rete. Infatti una mutazione nella molecola p53 determina la disattivazione della proteina, che trasforma la cellula in una cellula tumorale, che può portare anche alla morte dell'organismo. Conoscere la rete p53, quindi, è un passo importante per studiare una terapia farmacologica contro i tumori.

Quindi guardare alla cellula come una rete interconnessa, e non più come un semplice contenitore di sostanze chimiche indipendenti, ci consentirà di sviluppare metodi diagnostici e terapeutici migliori, sempre più individualizzati (Barabási, 2004).

3. Gli organismi coinvolti nell'analisi e il concetto di simbiosi

In questo studio sono state analizzate le reti metaboliche di 18 microrganismi. Si è partiti dall'analisi della rete di *Wolbachia*, endosimbionte del nematode filaride *Brugia malayi*. Questo nematode è l'agente infettivo dell'elefantiasi, che colpisce milioni di persone nelle zone tropicali e sub-tropicali. *Wolbachia* è un batterio mutualista che gioca un ruolo nella biologia degli ospiti e che partecipa alla patogenesi delle filariosi (Brattig et al., 2004). Incrementare le conoscenze su questi microrganismi e il loro rapporto con gli ospiti potrebbe quindi permettere di comprendere meglio le patologie e sviluppare nuovi metodi per il controllo delle infezioni e cure più efficaci.

A scopo comparativo sono state poi analizzate le reti metaboliche di altri batteri, alcuni appartenenti all'ordine Rickettsiales, lo stesso cui appartiene *Wolbachia*, altri simbionti mutualisti di artropodi, *Wigglesworthia glossinidia*, *Sodalis glossinidius* e *Buchnera aphidicola*, e quattro batteri a vita libera, *Escherichia coli*, *Salmonella enterica*, *Agrobacterium tumefaciens* e *Magnetospirillum magneticum*. Lo scopo di questi confronti è quello di verificare la presenza di eventuali differenze nelle reti metaboliche degli organismi in studio dovute alle loro relazioni filogenetiche o al diverso stile di vita.

1. CONCETTO DI SIMBIOSI

Il termine simbiosi deriva dal greco e, letteralmente, significa “vivere con”: descrive la relazione nella quale organismi (di specie diverse nell'accezione più condivisa) vivono insieme. Il termine è stato coniato nel 1879 dal patologo vegetale Anton De Bary, per descrivere i licheni, una simbiosi tra un'alga e un fungo. Anche se in seguito il termine simbiosi è stato utilizzato, scorrettamente, come sinonimo di mutualismo, De Bary gli attribuì un significato più vasto, includendo parassitismo, mutualismo e commensalismo.

Si possono infatti distinguere diverse categorie di relazioni interspecifiche:

- **Commensalismo:** relazione in cui uno dei due organismi (commensale) viene avvantaggiato dalla relazione stessa, mentre l'altro (ospite) non ne viene influenzato. Si distinguono diverse forme di commensalismo:

- Inquilinismo: l'ospite fornisce riparo al commensale;
- Amensalismo: contrario del commensalismo; l'amensale non trae vantaggio dalla relazione, mentre l'altro organismo viene svantaggiato o inibito;
- Foresia: il simbionte trae vantaggio facendosi trasportare dall'ospite.
- Mutualismo: relazione in cui entrambi gli organismi (mutualisti) hanno un vantaggio. Si distingue in mutualismo obbligato, in cui i mutualisti non potrebbero vivere in assenza dell'altro, e facoltativo, in cui gli organismi possono vivere anche separatamente.
- Parassitismo: relazione in cui un organismo (parassita) provoca un danno all'ospite sfruttandone le risorse o manipolando l'allocazione delle risorse dell'ospite stesso. Anche in questo caso si distingue tra parassitismo obbligato o facoltativo.

Anche se a livello teorico è possibile definire categorie ben distinte di simbiosi, le relazioni reali tra organismi spesso sono difficilmente riconducibili ad una sola tipologia. Infatti i confini tra i diversi tipi di simbiosi sono molto labili, e un simbionte può presentare caratteristiche che lo riconducono a più di una forma di relazione. Si devono quindi considerare mutualismo e parassitismo come estremi di un continuum di interazioni (Figura 7).

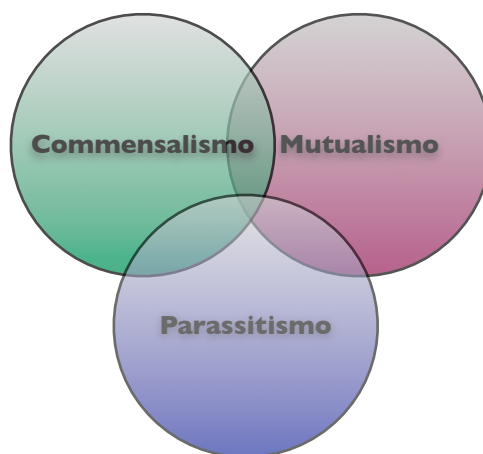


Fig. 7. Schema che mostra come il "confine" tra i diversi tipi di simbiosi possa essere sfumato e nelle zone di sovrapposizione risulta difficile definire il tipo di simbiosi.

1.1. L'endosimbiosi o simbiosi intracellulare

Quasi tutti gli organismi presi in esame (ad eccezione dei quattro a vita libera) sono batteri endosimbionti, o nel caso specifico più correttamente dei simbionti intracellulari, che vivono appunto all'interno delle cellule dell'ospite, spesso racchiusi in vacuoli o cellule specializzate, dette batteriociti. La simbiosi intracellulare è una relazione molto intima tra i due organismi, che conduce ad una elevata specializzazione e, in caso di mutualismo, ad una stretta interdipendenza tra essi. In questo caso la simbiosi diventa obbligatoria e nessuno dei due organismi è in grado di sopravvivere in assenza dell'altro. Il simbiote intracellulare giunge addirittura a delegare all'ospite la sua riproduzione, arrivando anche a manipolarla per massimizzare la sua diffusione. Infatti la trasmissione del simbiote avviene per via verticale, dalla madre alla prole, attraverso le uova o il nutrimento. Il batterio cercherà perciò di ridurre la sua virulenza in modo da facilitare la riproduzione dell'ospite infetto, e quindi favorire la propria diffusione. La presenza di una trasmissione verticale conduce spesso ad una congruenza nelle filogenesi di ospiti e simbionti.

Una caratteristica dei batteri endosimbionti è il fenomeno del *genome shrinkage*, cioè perdita di quei geni non più necessari alla vita all'interno dell'ambiente cellulare dell'ospite: di conseguenza il simbiote diventa dipendente dall'ospite per diversi aspetti. A sua volta, nel caso di relazioni mutualistiche, anche l'ospite può ricevere dal simbiote alcuni composti, come un'integrazione alla dieta nel caso delle simbiosi nutrizionali. Questo tipo di relazione coinvolge ospiti, soprattutto insetti, con una dieta sbilanciata, ad esempio a base di sangue, linfa vegetale o legno, carente di diverse sostanze, come amminoacidi e vitamine. Queste sostanze vengono fornite loro dai batteri endosimbionti, da cui l'ospite diventa dipendente.

Gli endosimbionti possono essere classificati in due categorie:

- i simbionti primari sono associati ai loro ospiti da molti milioni di anni, formano associazioni obbligate e mostrano coevoluzione con gli ospiti;
- i simbionti secondari mostrano una associazione sviluppatasi più recentemente, a volte sono trasmessi anche per via orizzontale e non vivono in batteriociti; le endosimbiosi secondarie non sono di tipo obbligato.

2. GLI ORGANISMI ANALIZZATI

Segue una breve trattazione dei batteri oggetto di studio, allo scopo di chiarire il tipo di simbiosi in cui sono coinvolti e le informazioni utili all'interpretazione delle reti metaboliche studiate.

2.1. *Wolbachia*

Gli appartenenti al genere *Wolbachia* sono batteri gram-negativi, simbionti intracellulari di artropodi (come insetti, acari, ragni, isopodi) e nematodi filaridi. *Wolbachia* appartiene all'ordine Rickettsiales, famiglia Anaplasmataceae, del gruppo α -proteobatteri. Nonostante ci sia un'unica specie valida, *Wolbachia pipientis*, vi è grande diversità molecolare tra i rappresentanti del genere nei diversi ospiti. Mancando una definizione formale di specie, si preferisce utilizzare il termine generico, *Wolbachia*, nel citare questi organismi, mentre nella filogenesi di *Wolbachia* vengono distinti diversi "supergruppi": A e B sono stati rinvenuti negli insetti (Werren et al., 1995), C e D nei nematodi filaridi (Bandi et al., 1998), E nei collemboli (Vandekerckove et al., 1999), F in artropodi (termiti) e alcuni nematodi filaridi (Lo et al., 2002). Inoltre negli ultimi anni sono stati descritti il supergruppo G, comprendente *Wolbachia* simbiote dei ragni (Rowley et al., 2004) e H, ritrovati in altre termiti, diverse da quelle del gruppo F (Berdestein e Rosengaus, 2005).

La scoperta di *Wolbachia* negli artropodi si deve a Hertig e Wolbach, che nel 1924 identificarono delle rickettsie negli ovari della zanzara *Culex pipiens*, che nel 1936 vennero chiamati *Wolbachia pipientis* da Hertig. Negli anni '50 in diversi studi si osservò che alcuni incroci tra popolazioni diverse di zanzare non avevano successo, a causa di un fattore con eredità citoplasmatica: si parlò quindi di incompatibilità citoplasmatica. Solo negli anni '70 questo fatto venne collegato alla presenza di *Wolbachia*: Yen e Barr scoprirono infatti che il fattore citoplasmatico responsabile dell'incompatibilità era costituito dal batterio.

La scoperta del batterio nelle filarie è invece più recente: negli anni '70 Harada et al. hanno osservato al microscopio elettronico la presenza di corpi tondeggianti nei tessuti della filaria *Dirofilaria immitis*. In seguito queste strutture vennero identificate come batteri e solo negli anni '90 si riconobbero come strettamente affini a *Wolbachia*. Nelle filarie, il batterio è stato individuato solo in due sottofamiglie, Onchocercinae e Dirofilarinae, della famiglia delle Onchocercidae. Una particolarità

di *Wolbachia* nelle filarie è che, nelle specie in cui il simbionte è presente, si registra una prevalenza del 100%, mentre ci sono specie che mancano totalmente dei simbionti (Casiraghi et al., 2004).

Wolbachia vive all'interno delle cellule dell'ospite, in un vacuolo circondato da tre strati di membrane: lo strato più esterno ha origine dall'ospite, seguito dalla parete cellulare del batterio e dalla sua membrana plasmatica. La localizzazione dei batteri varia con l'ospite: negli artropodi si trovano nelle cellule degli organi riproduttivi, ovari (nelle cellule nutrici) e testicoli, nelle cellule dei tubuli malpighiani e in altri distretti corporei. Nelle filarie si trovano invece nelle corde laterali in entrambi i sessi e, nelle femmine, negli ovari e negli embrioni in via di sviluppo.

Sia negli artropodi che nei nematodi, *Wolbachia* è trasmessa per via verticale, dalla madre alla prole, anche se negli artropodi ci sono evidenze per una trasmissione anche orizzontale. Questa differenza potrebbe essere riconducibile al diverso ruolo del batterio nei due tipi di ospite: di solito, infatti, una trasmissione orizzontale è legata ad un rapporto di tipo parassitario. La trasmissione verticale, invece, porta ad una riduzione della virulenza, e ad un tentativo del batterio di aumentare la capacità riproduttiva dell'ospite e, con questa, la sua possibilità di trasmissione (Bandi et al., 2001). In effetti, mentre negli artropodi *Wolbachia* si comporta da parassita riproduttivo, nelle filarie ci sono evidenze per un'interazione mutualistica.

2.1.1. *Wolbachia* simbionte di *Brugia malayi* (*wBm*)

Nelle filarie *Wolbachia* sembra avere un ruolo positivo, a differenza di quanto accade negli artropodi. Ci sono diverse evidenze a supporto di questa affermazione:

1. La distribuzione di *Wolbachia*, che è fissata nelle specie infette (100% di prevalenza);
2. La congruenza tra le filogenesi dell'ospite e del simbionte: questo è dovuto anche alla trasmissione strettamente verticale del simbionte;
3. Il trattamento con antibiotici in grado di eliminare *Wolbachia* ha determinato effetti negativi anche sulle filarie, quali il blocco dell'embriogenesi e il blocco dello sviluppo degli stadi giovanili: questo porterebbe a pensare che la presenza del simbionte favorirebbe la riproduzione e lo sviluppo dell'ospite (Rajan, 2004);
4. *Wolbachia* possiede una proteina di superficie (WSP) che media il rapporto con l'ospite; in *wMel* è stata riscontrata una selezione positiva per questa proteina,

che consente al batterio di rispondere alle strategie di difesa messe in atto dall'ospite, cioè a variazioni nel "recettore" di *wsp*. In *Wolbachia* nelle filarie non è invece stata osservata selezione positiva per questa proteina (Jiggins et al., 2002; Baldo et al., 2002);

5. Il sequenziamento dei genomi dei simbionti ha evidenziato le ridotte dimensioni del genoma di *wBm* (1,08 Mb) in confronto a quello di *wMel* (1,27 Mb) (Foster et al., 2005; Wu et al., 2004): similmente ad altri mutualisti, come *Buchnera aphidicola* e *Wigglesworthia glossinidia*, anche *wBm* avrebbe subito processi di genome shrinkage, in modo maggiore del parassita *wMel*.

Il sequenziamento del genoma di *wBm* ha permesso di analizzare le capacità metaboliche di questo batterio e i possibili scambi che questo ha con l'ospite. Infatti uno dei ruoli principali di *Wolbachia* nella simbiosi con la filaria è quello di fornire all'ospite diversi metaboliti. Mentre i pathway per la sintesi degli amminoacidi sono ridotti ad uno soltanto, sono presenti pathway completi per la sintesi di purine e pirimidine, e sembra che i nucleotidi vengano forniti anche all'ospite. Inoltre nel genoma di *wBm* sono presenti tutti i geni per gli enzimi per la biosintesi di riboflavina e FAD, e anche questi coenzimi potrebbero essere trasferiti all'ospite. Infine, un importante elemento che viene fornito all'ospite da *Wolbachia* è l'eme: questo potrebbe svolgere un ruolo importante nella riproduzione e sviluppo della filaria, e potrebbe essere l'interruzione del trasporto dell'eme la causa degli effetti negativi sulle filarie dovuti al trattamento con antibiotici. Il simbionte dipenderebbe invece dall'ospite per altri cofattori e vitamine, quali coenzima A, NAD, biotina, ubiquinone e altri. Inoltre i principali substrati di crescita per il batterio, zuccheri e amminoacidi, vengono prelevati dall'ambiente extracellulare, quindi dal nematode (Foster et al., 2005).

2.1.2. *Wolbachia* simbiote di *Drosophila melanogaster* (*wMel*)

L'effetto della presenza di *Wolbachia* nelle cellule dell'ospite artropode è un'alterazione del processo riproduttivo, perciò *Wolbachia* è definito parassita riproduttivo. Attualmente i molti studi condotti su *Wolbachia* hanno portato ad ampliare le conoscenze su questo simbionte, che ha mostrato di avere molti effetti sugli ospiti: oltre all'incompatibilità citoplasmatica, infatti, è responsabile anche di induzione della partenogenesi (Stouthamer et al., 1990), femminizzazione di maschi genetici (Rousset et al., 1992), male-killing (Hurst et al., 1996). Inoltre, sono stati riscontrati

anche casi di ruoli positivi del batterio, con accrescimento della capacità riproduttiva degli ospiti infetti (Girin e Bouletreau, 1995).

Incompatibilità citoplasmatica

L'incompatibilità citoplasmatica indotta da *Wolbachia* negli artropodi viene distinta in due classi:

- nell'incompatibilità citoplasmatica unidirezionale si verifica incompatibilità tra spermatozoi provenienti da un maschio infetto e uova non infette: questo determina la morte degli embrioni o, nel caso di aplodiploidia, spostamento del rapporto tra i sessi a favore dei maschi. Nel caso opposto (maschio non infetto, uova infette) l'incrocio è compatibile.
- nell'incompatibilità citoplasmatica bidirezionale si osserva l'incompatibilità riproduttiva tra artropodi infetti da diversi ceppi di batteri. Questo effetto è molto importante dal punto di vista evolutivo: se popolazioni della stessa specie ospitano batteri di ceppi diversi potrebbe insorgere una barriera riproduttiva che a lungo termine porterebbe ad una speciazione.

Non è ancora del tutto noto il meccanismo alla base dell'incompatibilità citoplasmatica, ma è stato proposto un modello per spiegarlo, il modello modification-rescue (Werren, 1997). Secondo questo modello *Wolbachia* produrrebbe nel maschio un fattore che modifica lo spermatozoo (mod⁺), che quindi non sarà in grado di far sviluppare lo zigote. Se però la femmina è infetta (res⁺), quindi l'uovo contiene il batterio, questo produrrà un fattore rescue, di salvataggio, che permette di compensare la modifica dello spermatozoo.

Quindi si possono avere diverse situazioni:

		femmina	
		Res ⁺	Res ⁻
maschio	Mod ⁺	+	-
	Mod ⁻	+	+

Dal punto di vista citologico, si è visto che la presenza di *Wolbachia* nel maschio modifica il processo di sviluppo degli spermatozoi, per cui il fattore di modificazione impedirebbe alla cromatina di decompattarsi dopo la fecondazione. Infatti negli incroci incompatibili solo il pronucleo femminile forma cromosomi individuali, mentre la cromatina paterna non si decompatta e viene quindi eliminata. Questo

porta alla formazione di embrioni aploidi che, nelle specie diploidi vengono eliminati mentre nelle specie aplodiploidi si svilupperanno in individui maschi.

Induzione di partenogenesi

Sembra che questo effetto di *Wolbachia* sull'ospite sia legato al tentativo del batterio di favorire la sua trasmissione. Infatti, dato che solo le femmine trasmettono i simbionti ai gameti, per il batterio è più favorevole un rapporto sessi spostato verso le femmine (Stouthamer et al., 1999). Questo effetto è stato rilevato in ospiti appartenenti all'ordine degli imenotteri. Il meccanismo con il quale *Wolbachia* provoca la partenogenesi consiste nella modificazione della prima divisione mitotica nello zigote. Nelle uova infette la divisione viene bloccata all'anafase, per cui si forma un nucleo diploide in un uovo fecondato: dato che gli ospiti imenotteri sono aplodiploidi, queste uova si svilupperanno in individui femmina. Se la popolazione di femmine viene trattata con antibiotici i maschi riappaiono, a dimostrare che a causare la partenogenesi è la presenza del batterio.

Femminizzazione di maschi genetici

Questo effetto è stato osservato in alcuni isopodi (*Armadillum vulgare*) e lepidotteri. Negli isopodi, in particolare, si sa che la determinazione del sesso è legata alla secrezione di un ormone da parte della ghiandola androgenica durante lo sviluppo. Se non avviene la secrezione dell'ormone, gli individui si sviluppano come femmine. Sembra che *Wolbachia* intervenga a livello della ricezione del segnale ormonale, trasformando così i maschi genetici in femmine.

Effetto male-killing

Questo effetto sembra essere abbastanza diffuso in natura: in più di 20 specie di insetti è stato riscontrato un fattore citoplasmatico che uccide gli embrioni maschi, e in tutti questi casi il fenomeno è curabile con antibiotici, quindi è dovuto a batteri. Tra i batteri responsabili di questo effetto è stato trovato anche *Wolbachia*, anche se il meccanismo con cui viene prodotto non è ancora noto.

Effetti positivi di *Wolbachia*

In alcuni casi sono stati riscontrati effetti positivi della presenza di *Wolbachia* nelle cellule di artropodi, per cui viene definito mutualista. Ad esempio in una vespa parassitoide, *Asobara tabida*, le femmine portatrici del batterio hanno una maggiore capacità riproduttiva rispetto alle altre; il trattamento con antibiotici determina l'incapacità delle femmine a produrre oociti maturi. In un altro caso, quello del co-

leottero *Tribolium confusum*, la presenza del batterio aumenta la fecondità dei maschi.

Parassitismo

In *Drosophila melanogaster* è stato anche osservato un ceppo di *Wolbachia pipientis* che si comporta come un vero parassita: dopo lo sviluppo, nell'adulto la presenza del batterio provoca una degenerazione dei tessuti che porta alla morte precoce dell'individuo. Questo ceppo è stato chiamato "popcorn strain".

Nonostante si conoscano i numerosi effetti fenotipici di *Wolbachia* sull'ospite, i meccanismi responsabili di questi effetti non sono ancora del tutto noti. Il sequenziamento del genoma dei batteri simbiotici di *Drosophila melanogaster* (Wu et al., 2004) ha permesso di identificare alcuni fattori che interverrebbero nella relazione tra *Wolbachia* e l'ospite. In particolare si è scoperto che *Wolbachia* possiede un sistema di secrezione di Tipo IV, che in molti batteri veicola sostanze implicate nella virulenza, ad esempio il plasmide Ti in *Agrobacterium tumefaciens*. In *Wolbachia* potrebbe essere implicato nello stabilirsi e mantenersi dell'infezione e possibilmente anche nel provocare diversi fenotipi riproduttivi (Wu et al., 2004). Dal punto di vista metabolico, *wMel* ha ridotte capacità di trasporto, di utilizzo dei substrati e di biosintesi di intermedi metabolici. Dall'analisi del genoma è emerso che gli unici trasportatori sono per amminoacidi e, essendoci pathway per il metabolismo di alcuni amminoacidi e un limitato metabolismo dei carboidrati, si pensa che questi batteri ottengano la maggior parte della loro energia dagli amminoacidi.

2.2. Rickettsiales

All'ordine dei Rickettsiales, oltre a *Wolbachia*, appartengono altri generi di batteri endosimbionti, perlopiù parassiti. Di questi sono state analizzate alcune specie del genere *Rickettsia*, *Ehrlichia canis* e due specie del genere *Anaplasma*.

2.2.1. *Rickettsia* spp.

I batteri del genere *Rickettsia* sono responsabili di diverse patologie nell'uomo. Le specie appartenenti a questo genere vengono distinte, dal punto di vista filogenetico, in tre gruppi, che corrispondono ai tipi di malattia provocata nell'uomo:

- il gruppo del tifo (TG), che comprende *R. prowazekii* e *R. typhi*.
- il gruppo della "spotted fever" (SFG), comprendente, tra le altre, *R. conorii* e *R. felis*.

- la specie *R. bellii* dal punto di vista filogenetico rimane distante dalle altre (Ogata et al., 2006).

Le rickettsiae sono parassiti intracellulari dixeni (i.e. caratterizzate da un ciclo con due ospiti), un artropode e un mammifero. In particolare gli artropodi ematofagi (zecche, pidocchi e pulci) funzionano come vettori dei parassiti nell'ospite mammifero.

Le rickettsiae vengono inoculate nel derma dell'ospite dalla puntura di una zecca o attraverso la pelle danneggiata dalle feci di pidocchi o pulci. I batteri poi si diffondono attraverso il sistema sanguigno e infettano l'endotelio, dove aderiscono alle cellule ospiti. Dopo l'adesione, inducono la fagocitosi nella cellula ospite, e quindi si diffondono nel citoplasma.

In seguito, le rickettsie responsabili del tifo si moltiplicano, provocando la lisi della cellula ospite, da cui poi escono. Gli appartenenti al gruppo della *spotted fever*, invece, non lisano la cellula ospite ma escono da essa stimolando la polimerizzazione delle code di actina dell'ospite, che li muovono verso la membrana plasmatica, da cui emergono.

Rickettsia prowazekii

Ospite: uomo

Tipo di simbiosi: è un parassita, responsabile del tifo epidemico nell'uomo.

Dimensione del genoma: 1.1 Mb

Numero di geni: 834, di cui 549 codificanti proteine

Anno di pubblicazione del genoma: 1998 (Andersson et al.)

Possibili passaggi di molecole tra ospite e parassita: all'inizio del ciclo infettivo *R. prowazekii* usa l'ATP dell'ospite, pur essendo in grado di produrla da sé (Andersson et al., 1998). Inoltre, mancando gli enzimi glicolitici, i batteri del genere *Rickettsia* potrebbero ottenere dall'ospite il glicerolo-3-fosfato (Dunning Hotopp, 2006).

Rickettsia bellii

Ospite: uomo; vettore: zecche

Tipo di simbiosi: parassita

Dimensione del genoma: 1.5 Mb

Numero di geni: 1429

Anno di pubblicazione del genoma: 2006 (Ogata et al.)

Rickettsia conorii

Ospite: uomo; vettore: zecca marrone del cane, *Rhipicephalus sanguineus*.

Tipo di simbiosi: è un parassita, responsabile della *spotted fever* mediterranea.

Dimensione del genoma: 1.3 Mb

Numero di geni: 1374

Anno di pubblicazione del genoma: 2001 (Ogata et al.)

Rickettsia felis

Ospite: uomo; vettori: le pulci *Ctenocephalides felis* e *Pulex irritans*

Tipo di simbiosi: è un parassita, responsabile della *flea-borne spotted fever*

Dimensione del genoma: 1.5 Mb

Numero di geni: 1512 codificanti proteine

Anno di pubblicazione del genoma: 2005 (Ogata et al.)

Possibili passaggi di molecole tra ospite e parassita: *R. felis* possiede un meccanismo specifico di comunicazione con l'ospite (ankirina).

Rickettsia typhi

Ospite: uomo; vettori: le pulci *Xenopsylla cheopis*, *Ctenocephalides felis* e *Polyplax spinulos*.

Tipo di simbiosi: è un parassita, responsabile del tifo endemico nell'uomo.

Dimensione del genoma: 1.1 Mb

Numero di geni: 877

Anno di pubblicazione del genoma: 2004 (McLeod et al.)

Possibili passaggi di molecole tra ospite e parassita: possiede un sistema di secrezione di Tipo IV; inoltre sono state rinvenute proteine responsabili di trasporti attraverso la membrana, come l'ADP/ATP traslocasi che permette di acquisire ATP dalla cellula ospite.

2.2.2. *Ehrlichia canis*

Ehrlichia canis è l'agente eziologico dell'erlichiosi canina. Il ciclo vitale di *E. canis* prevede la presenza di due ospiti: un ospite intermedio artropode, la zecca *Rhipi-*

cephalus sanguineus, e l'ospite definitivo mammifero, il cane, in cui risiede in vacuoli all'interno dei monociti e dei macrofagi.

Dimensione del genoma: 1.3 Mb

Numero di geni: 984

Anno di pubblicazione del genoma: 2006 (Mavromatis et al.)

Possibili passaggi di molecole tra ospite e parassita: dato che mancano gli enzimi per la sua produzione, sembra che il colesterolo venga acquisito dall'ospite, e quindi incorporato nella membrana esterna del batteri. Sono presenti, invece, pathway per la biosintesi di alcuni amminoacidi, purine, pirimidine, di lipidi e fosfolipidi (Mavromatis et al., 2006).

2.2.3. *Anaplasma* spp.

Anaplasma phagocytophilum

Ospite: uomo; vettori: pulci del genere *Isodex*.

Tipo di simbiosi: è un parassita, responsabile dell'anaplasmosi dei granulociti nell'uomo.

Dimensione del genoma: 1.5 Mb

Numero di geni: 747 geni con funzione assegnata

Anno di pubblicazione del genoma: 2006 (Dunning Hotopp et al.)

Possibili passaggi di molecole tra ospite e parassita: *A. phagocytophilum* è in grado di sintetizzare tutti i nucleotidi, cofattori e vitamine. Questo suggerisce che potrebbe fornire questi composti all'ospite. Ha invece ridotte capacità di sintesi di amminoacidi, che potrebbe acquisire dall'ospite (Dunning Hotopp et al., 2006).

Anaplasma marginale

Ospite: bovini; vettori: pulci del genere *Ixodes*.

Tipo di simbiosi: è un parassita, che invade gli eritrociti bovini.

Dimensione del genoma: 1.2 Mb

Numero di geni: 949 geni codificanti proteine

Anno di pubblicazione del genoma: 2005 (Brayton et al.)

Possibili passaggi di molecole tra ospite e parassita: il genoma di *A. marginale* possiede pochi geni codificanti enzimi per la biosintesi degli amminoacidi, anche se sono stati trovati trasportatori di queste molecole. Sono inoltre presenti trasportatori per ribonucleotidi e fosfato (Brayton et al., 2005).

2.3. Endosimbionti mutualisti di insetti

2.3.1. *Wigglesworthia glossinidia*

Wigglesworthia glossinidia è un batterio appartenente al gruppo dei γ -proteobatteri, simbionte primario della mosca tsetse (genere *Glossina*). È un simbionte intracellulare, localizzato in batteriociti, formati da cellule epiteliali dell'ospite modificate. La sua trasmissione avviene per via verticale: le uova delle mosche tsetse si sviluppano all'interno dell'utero materno, e vengono deposte 6-8 larve di terzo stadio. I nutrienti, e con questi i simbionti intracellulari, vengono forniti alle larve in sviluppo nell'utero attraverso la secrezione di una ghiandola. Non è ancora noto se il simbionte sia trasmesso nei batteriociti o come batteri singoli. Analizzando però il genoma di *W. glossinidia* si è visto che esso ha mantenuto i geni per la codifica dell'apparato flagellare. Dato che non sono stati rinvenuti nei batteriociti batteri provvisti di flagelli o dotati di mobilità, si pensa che questo apparato sia implicato nella loro trasmissione (Akman et al., 2002).

La simbiosi che coinvolge la mosca tsetse e questo batterio è di tipo nutrizionale: *W. glossinidia* integra la dieta dell'insetto (composta da sangue di vertebrati) fornendogli vitamine essenziali, in particolare la vitamina B. Da parte sua, il batterio può riprodursi in condizioni controllate e ricche di nutrienti, nelle cellule specializzate dell'ospite. Studi di manipolazione in cui *W. glossinidia* è stato eliminato attraverso antibiotici hanno mostrato una riduzione della fertilità, la riduzione del 50% delle pupae nate dalle mosche trattate con antibiotico, e una minore longevità nella generazione F₁ (Dale e Welburn, 2001).

La congruenza tra la filogenesi di *W. glossinidia* e quella dell'ospite dimostrerebbe l'antichità di questa simbiosi, che sembra risalire a 50-80 milioni di anni fa (Chen et al., 1999). Questo sarebbe dimostrato anche dalla dimensione del genoma del batterio, che ha subito eventi di *genome shrinkage*. Il genoma di *W. glossinidia*, infatti, è composto da sole 700 kb e codifica per 621 prodotti. A dimostrazione del fatto che si tratta di un simbionte nutrizionale, *W. glossinidia* ha mantenuto 62 geni coinvolti nella biosintesi di cofattori, gruppi prostetici e carriers. Di particolare interesse, inoltre, è il fatto che il genoma di *W. glossinidia* ha mantenuto i geni per la costruzione della parete cellulare e la biosintesi di lipopolisaccaridi e peptidoglicani (Akman et al., 2002). La presenza di queste strutture, insieme al flagello, sono caratteristiche tipiche dei parassiti, in quanto legate all'invasione della cellula ospite, mentre

nel caso dei mutualisti queste vengono spesso perse. Il fatto che *W. glossinidia* possieda la capacità di svilupparle la colloca, se si osservasse puramente il suo genoma, a metà strada tra un parassita e un mutualista.

2.3.2. *Sodalis glossinidius*

Sodalis glossinidius è anch'esso un simbiote della mosca tsetse, appartenente al gruppo dei γ -proteobatteri. Si localizza sia a livello intracellulare che extracellulare in diversi tessuti, quali l'intestino medio, i muscoli, il corpo grasso e le ghiandole salivari. La sua trasmissione è verticale e, come *Wigglesworthia glossinidia*, raggiunge le larve intrauterine attraverso la ghiandola del latte. Si tratta di un simbiote secondario che si sarebbe separato da un antenato a vita libera in tempi recenti: questo sarebbe dimostrato dalla possibilità di coltivare il batterio in un mezzo privo di cellule (Beard et al., 1993), dal fatto che condivide molti geni (più di 1800) con *E. coli* (Akman et al., 2001), e dalle dimensioni del suo genoma, circa 4 Mb, elevata per un simbiote (Wernegreen, 2002). Il sequenziamento del genoma di *S. glossinidius* (Toh et al., 2006), ha mostrato che questo batterio possiede caratteristiche che lo includerebbero in una categoria intermedia, di transizione tra un batterio a vita libera e un mutualista. Analisi filogenetiche hanno infatti mostrato che questo batterio è strettamente imparentato con i generi *Salmonella*, *Yersinia* e *Photobacterium*, oltre ad *E. coli* K-12. A differenza di altri simbiotici, inoltre, come *W. glossinidia* e *Buchnera aphidicola*, che hanno subito processi di riduzione del genoma in seguito all'instaurarsi della simbiosi, il genoma di *S. glossinidius* è più simile a quello di batteri a vita libera. Sta però subendo un processo per cui alcuni pathways non più necessari per la vita all'interno dell'ospite vengono disattivati. Dal punto di vista metabolico, *S. glossinidius* ha mantenuto la capacità di sintetizzare tutti gli amminoacidi, tranne l'alanina, mentre ha perso i geni per la degradazione degli amminoacidi e quelli coinvolti nella respirazione anaerobica, come le nitrito e fumarato reductasi, la glicero-3-fosfato deidrogenasi e la formiato deidrogenasi (Toh et al., 2006). Dal confronto con *W. glossinidia* è emerso che i geni di quest'ultimo sono condivisi anche da *S. glossinidius*, ad eccezione di quelli che codificano per i pathway biosintetici di tiamina, cobalamina e molidopterinina, che sarebbero i fattori che *W. glossinidia* fornisce all'ospite.

Anche *S. glossinidius* sembra avere un ruolo di mutualista nella simbiosi con la mosca tsetse. Infatti eliminando selettivamente *S. glossinidius* dalle mosche con trattamen-

to antibiotico (streptozotocina), si è ottenuta una minore longevità nelle mosche. La mancanza del simbiote secondario, inoltre, ha provocato un ridotto tasso di infezione da parte dei tripanosomi, parassiti responsabili della malattia del sonno, di cui le mosche tsetse sono vettori (Dale e Welburn, 2001). Infatti il tripanosoma, per infettare la mosca, deve evadere il sistema immunitario dell'ospite, che agisce attraverso la lectina, secreta durante il pasto della mosca. *S. glossinidius* è in grado di produrre un inibitore della lectina, l'N-acetil-D-glucosammina, e in questo modo promuove l'infezione da parte del tripanosoma. Questo aspetto della simbiosi è molto importante dal punto di vista applicativo, in quanto la rimozione di *S. glossinidius* riduce la capacità di trasmissione del tripanosoma da parte della mosca tsetse. Di conseguenza si stanno studiando possibili applicazioni per il controllo della diffusione del tripanosoma.

2.3.3. *Buchnera aphidicola*

Buchnera aphidicola, appartenente al gruppo dei γ -proteobatteri, è un simbiote nutrizionale degli afidi. Questi insetti si nutrono di linfa, sostanza carente di amminoacidi, che vengono loro forniti dal simbiote. La simbiosi è di tipo obbligatorio, con i batteri che sono inclusi in batteriociti presenti nell'emocele. La trasmissione del simbiote è di tipo verticale, attraverso le uova e gli embrioni.

Oggi sono studiati tre ceppi di *B. aphidicola*, distinti in base all'ospite: *Buchnera Baisongia pistaciae* (BBp), *Buchnera Acyrtosiphon pisum* (BAp) e *Buchnera Schizaphis graminum* (BSg). Dall'analisi del genoma dei tre ceppi è emerso che il genere si è separato circa 200-250 milioni di anni fa dall'antenato comune simbiote; da questo antenato, la linea di BBp si è separata 80-150 milioni di anni fa dall'antenato comune degli altri due ceppi, che si sono divisi circa 50-70 milioni di anni fa (van Ham et al., 2002; Tamas et al., 2002). I genomi di questi batteri hanno una dimensione di circa 618, 652 e 653 kb rispettivamente, con 638 geni totali, 544 in BBp, 564 in BAp e 545 in BSg. Si tratta complessivamente di genomi molto piccoli, vicini al minimo genoma di un essere vivente conducente vita libera, quello di *Mycoplasma genitalium* (580 kb) e sette volte più piccolo del genoma di *E. coli*, con cui è imparentato. Questa caratteristica è dovuta al fatto che i genomi di *B. aphidicola* hanno subito un processo di riduzione che l'hanno ridotto di almeno 2000 geni dal momento della separazione della sua linea da quella di *E. coli* (Tamas et al., 2002). In particolare, i gruppi funzionali che hanno subito maggiori riduzioni sono quelli di trasporto e regolazione,

mentre la traduzione e il metabolismo degli amminoacidi hanno subito minore *shrinkage*. Nel genoma di *B. aphidicola* sono stati individuati 54 geni coinvolti nella biosintesi degli amminoacidi essenziali per gli ospiti, mentre quelli per gli amminoacidi non essenziali non sono presenti, ad indicare che il passaggio di sostanze tra ospite e simbionte è bidirezionale (Shigenobu et al., 2000). Nonostante sia provata l'esistenza di un flusso di sostanze tra i due organismi, *B. aphidicola* ha un ridotto apparato di trasporto, con un solo gene che codifica per un trasportatore specifico, quello per l'importazione di mannitolo e glucosio. Un altro aspetto interessante riguarda la riduzione dei composti dello zolfo: gli afidi non sono in grado di ridurre il solfato, mentre *B. aphidicola* possiede i geni per questo pathway e inoltre vari esperimenti hanno mostrato che il sistema *Buchnera*-batteriocita è responsabile dell'assimilazione del solfato (Douglas et al., 1988). Altre riduzioni hanno coinvolto i geni per la ricombinazione e riparazione del DNA; inoltre mancano i geni per la produzione di lipopolisaccaridi, quindi la superficie cellulare di *B. aphidicola* è vulnerabile, anche se la presenza di strutture esterne è legata più ad un rapporto di tipo parassita-ospite che non ad un mutualismo. La presenza di questo genoma minimo e il fatto che *B. aphidicola* non sia in grado di vivere al di fuori della cellula ospite conduce a considerare questo batterio come un organello cellulare nascente.

2.4. Batteri a vita libera

2.4.1. *Agrobacterium tumefaciens*

Agrobacterium tumefaciens appartiene al gruppo degli α -proteobatteri, famiglia Rhizobiaceae. È un microrganismo che vive nel suolo, ed è responsabile dell'insorgenza del tumore del colletto nelle piante. La patogenicità di questo batterio è dovuta alla presenza, all'interno della cellula, di un plasmide, chiamato plasmide Ti (tumor inducer). Quando *A. tumefaciens* infetta le piante, inietta una regione specifica del plasmide, detta T-DNA, all'interno delle cellule vegetali lesionate e il T-DNA viene incorporato nel genoma della cellula ospite. Questo frammento contiene geni per la biosintesi di ormoni di crescita vegetali, oltre a geni per le opine, che servono per il sostentamento del batterio stesso. Gli ormoni provocano una sregolata proliferazione delle cellule a formare la galla.

Dimensione del genoma: 5,67 Mb

Numero di geni: 5419 geni codificanti proteine

Anno di pubblicazione del genoma: 2001 (Wood et al.; Goodner et al.).

2.4.2. *Magnetospirillum magneticum*

Magnetospirillum magneticum è un α -proteobatterio che fa parte del gruppo dei batteri magnetotattici. Questi microrganismi si trovano solitamente nella zona di transizione acqua-sedimenti. La proprietà distintiva dei batteri magnetici, è quella di acquisire ferro dal suolo, e convertirlo in particelle di magnetite o grigite. Queste particelle vengono accumulate in strutture intracellulari, i magnetosomi, dove formano delle catene. La presenza delle particelle magnetiche porta i batteri a orientarsi in direzione del campo magnetico terrestre.

Dimensione del genoma: 4,97 Mb

Numero di geni: 2290 geni codificanti proteine

Anno di pubblicazione del genoma: 2005 (Matsunaga et al.).

2.4.3. *Salmonella enterica*

I batteri del genere *Salmonella* appartengono alla famiglia Enterobacteriaceae del gruppo dei γ -proteobatteri. A questo genere appartengono due specie, *S. enterica* e *S. bongori*, suddivise in 7 sottospecie e 2400 serovar. Si tratta di batteri patogeni per l'uomo e gli animali. Nell'uomo sono causa di due patologie infettive, la febbre tifoide e le salmonellosi minori, trasmesse attraverso l'ingestione di cibi o bevande contaminate. Di *S. enterica*, in particolare, esistono sei sottospecie, suddivise a loro volta in migliaia di serovar. Quella presa in esame per questa indagine è *S. enterica enterica* serovar Choleraesuis. Ha un limitato spettro d'ospiti, e infetta occasionalmente anche l'uomo, in cui è causa di aneurismi infettivi.

Dimensione del genoma: 4,7 Mb

Numero di geni: 4445

Anno di pubblicazione del genoma: 2005 (Chiu et al.).

2.4.4. *Escherichia coli*

Escherichia coli è un γ -proteobatterio che vive nella parte inferiore dell'intestino di animali a sangue caldo, dove svolge un ruolo importante nel processo digestivo. Alcuni ceppi possono però provocare patologie, come malattie intestinali, infezioni del tratto urinario, meningite, peritonite, setticemia e polmonite. Il ceppo analizzato in questo lavoro, *E. coli* 536, è un patogeno extraintestinale, responsabile di infezioni del tratto urinario.

Dimensione del genoma: 4,9 Mb

Numero di geni: 4747

Anno di pubblicazione del genoma: 2006 (Brzuszkiewicz et al.).

In tabella 2 vengono riassunte, per i batteri studiati, le informazioni utili nell'analisi e confronto delle loro reti metaboliche.

Tab. 2. Elenco dei batteri presi in esame e caratteristiche generali del genoma e della simbiosi.

Organismo	Dimensioni genoma	Numero geni	Anno pubblicazione	Simbiosi
<i>Wolbachia pipientis</i> W _{Bm}	1080 kb	1218	2005	<i>Mutualista</i>
<i>Wigglesworthia glossinidia</i>	697 kb	621	2002	<i>Simbionte nutrizionale</i>
<i>Sodalis gossinidius</i>	4171 kb	2516	2005	<i>Commensale</i>
<i>Buchnera aphidicola</i> Ap	640 kb	574	2000	<i>Simbionte nutrizionale</i>
<i>B. aphidicola</i> Bp	615 kb	504	2003	<i>Simbionte nutrizionale</i>
<i>B. aphidicola</i> Sg	641 kb	545	2002	<i>Simbionte nutrizionale</i>
<i>Rickettsia bellii</i>	1522 kb	1429	2006	<i>Parassita</i>
<i>R. conorii</i>	1268 kb	1374	2001	<i>Parassita</i>
<i>R. felis</i>	1485 kb	1512	2005	<i>Parassita</i>
<i>R. prowazekii</i>	1111 kb	834	1998	<i>Parassita</i>
<i>R. typhi</i>	1111 kb	838	2004	<i>Parassita</i>
<i>Ehrlichia canis</i>	1315 kb	942	2006	<i>Parassita</i>
<i>Anaplasma marginale</i>	1197 kb	1185	2005	<i>Parassita</i>
<i>A. phagocytophilum</i>	1471 kb	1409	2005	<i>Parassita</i>
<i>Escherichia coli</i>	4938 kb	4747	2006	<i>Vita libera</i>
<i>Salmonella enterica</i>	4755 kb	4801	2005	<i>Vita libera</i>
<i>Magnetospirillum magneticum</i>	4967 kb	4824	2005	<i>Vita libera</i>
<i>Agrobacterium tumefaciens</i>	4915 kb	5419	2001	<i>Vita libera</i>

4. Metodi

1. COSTRUZIONE DELLE RETI METABOLICHE

1.1. Il database

Per l'analisi del metabolismo cellulare è stata utilizzata la sezione "EMP Metabolic Reconstruction" del database Puma2 (<http://compbio.mcs.anl.gov/puma2>). Questo database permette di consultare via internet le ricostruzioni dei pathway metabolici di decine di organismi. Ogni metabolismo riportato deriva da ricerche mirate sull'organismo, e dall'ipotesi di similarità agli organismi modello, nel caso di vie metaboliche ancora sconosciute.

È stato scritto un programma, sviluppato in Perl in ambiente WIN (Comandatore, unpublished) che consulta il database Puma2 direttamente da internet. Questo programma, dopo aver attuato un download di tutte le pagine HTML relative all'organismo, riconosce in ognuna di queste le reazioni metaboliche e le trascrive in un file TXT. Una volta scelto l'organismo di cui si vuole studiare il metabolismo è quindi possibile acquisirne tutte le reazioni oggi conosciute.

I dati scaricati dal sito del database sono stati esaminati e corretti: nel caso in cui il nome dei substrati fosse scritto in modi diversi (ad es. Glucose_6-phosphate e Glucose-6-phosphate) questi sono stati corretti in modo che ogni molecola avesse un nome univoco e fosse quindi rappresentata con un solo nodo nella rete.

1.2. Costruzione della rete metabolica

Una volta ottenuta la lista delle reazioni metaboliche per ogni organismo, un secondo programma, scritto in Perl in ambiente WIN (Comandatore, unpublished), ha permesso di costruire la rete.

Per la costruzione della rete in ogni reazione si considerano tre tipi di molecole: reagenti, prodotti e centro di reazione. Quest'ultimo può essere formato da uno o più enzimi, che concorrono a trasformare i reagenti in prodotti. L'utilizzo dei centri reazione, al posto dei singoli enzimi, per costruire la rete, è stata una scelta obbligata, dovuta alla struttura del database Puma2.

I link che uniscono i nodi della rete sono direzionali e seguono un criterio temporale. Ogni reagente è collegato al proprio centro di reazione attraverso una freccia di-

retta verso quest'ultimo. Allo stesso modo, ogni centro di reazione è unito ai propri prodotti con frecce dirette verso loro (Figura 8).

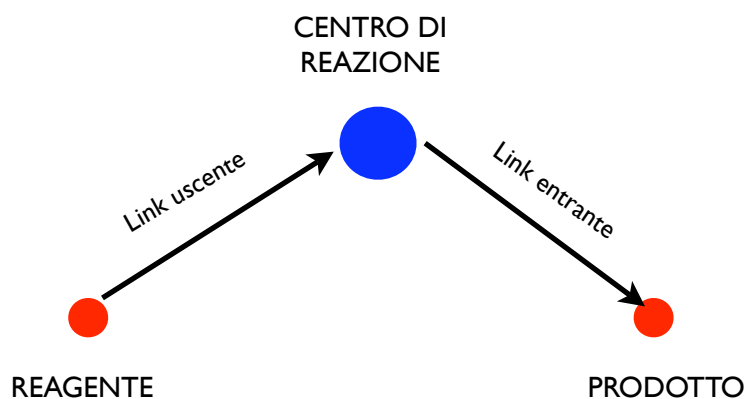


Fig. 8. Esempio di come è stata costruita la rete metabolica.

Il programma che è stato utilizzato in seguito per analizzare le reti è Pajek, sviluppato da Vladimir Batagelj e Andrej Mrvar (1998). Il file di rete (.net) accettato da questo programma è composto da due parti:

- un elenco di tutti i nodi, con associato un numero;
- una serie di coppie di numeri. Queste indicano l'esistenza di un link direzionale che unisce i due nodi a cui sono associati i numeri in questione, con direzione che va da quello a sinistra a quello a destra.

Il programma utilizzato per costruire le reti riporta, quindi, come risultato (oltre all'elenco dei nodi) una serie di coppie numeriche, formate dai numeri associati a:

- reagente- centro di reazione
- centro di reazione- prodotto.

2. ANALISI DELLE RETI METABOLICHE

Le reti ottenute sono state rappresentate graficamente utilizzando il programma Pajek. Per ognuna sono stati misurati diversi parametri necessari a descrivere la loro struttura.

2.1. Distribuzione dei gradi dei nodi

Dal momento che le reti metaboliche ottenute sono orientate, possono essere distinti due tipi di link. Vi sono link entranti in un nodo, cioè in un prodotto della reazione biochimica, e link uscenti da un nodo, che rappresenta un reagente.

I link uscenti ed entranti sono stati perciò analizzati separatamente. Per ogni nodo è stato calcolato il grado in entrata (k_{in}), cioè il numero di link che puntano a quel nodo e il grado in uscita (k_{out}), il numero di link uscenti da quel nodo. Per ogni valore di k_{in} e k_{out} è stata quindi calcolata la frequenza, e riportata in grafico. Il grafico in scala logaritmica mostra che la curva ottenuta segue la legge di potenza: sono stati quindi calcolati gli esponenti di grado (γ_{in} e γ_{out}) per ogni rete.

2.2. Diametro e cammino medio tra i nodi

Il diametro è stato calcolato come la maggiore delle distanze tra tutte le coppie di nodi della rete. Per distanza media tra i nodi si intende invece il valore medio delle distanze misurate tra tutte le coppie di nodi. Il calcolo di entrambi i valori è stato effettuato mediante il programma Pajek. Un ulteriore dato calcolato è la densità dei nodi della rete, ottenuta dal rapporto tra il numero totale di nodi e il diametro.

2.3. Hub

Anche per l'identificazione degli hub sono stati distinti i link entranti e uscenti da ogni nodo. Sono stati indicati come hub i nodi il cui grado, k_{in} e k_{out} , ecceda la media dei gradi di 3 volte la deviazione standard. Cioè: $k > \bar{k} + 3\sigma$. Dato che questa procedura ha consentito di individuare come hub molecole già note per il loro ruolo chiave nel metabolismo cellulare, si è deciso di includere nella lista altri nodi, con grado pari al limite imposto diminuito di 10. Infatti tra queste molecole potrebbero essere presenti substrati con un ruolo importante nel metabolismo ma che, a causa di imprecisioni nell'annotazione del genoma, risultano avere un grado inferiore a quello reale.

2.4. Effetti di attacchi ed errori sulla rete

Questa analisi è stata condotta sulla rete metabolica di *Wolbachia*. Per verificare se la rete metabolica mostra una tolleranza ad errori o attacchi mirati, questi sono stati simulati mediante l'eliminazione di nodi dal file che descrive la rete. In questo file, ogni nodo è rappresentato da un numero, e sono elencate le coppie di nodi connessi da un link nella rete. In particolare, per simulare un errore è stata generata una lista di numeri casuali con un generatore on-line (http://www.Regione.emilia-romagna.it/sin_info/generatore/random.asp). La lista generata è composta da

50 numeri, che costituiscono il 5% dei nodi della rete. Sono stati quindi eliminati dalla rete i nodi contrassegnati con questi numeri. Un attacco mirato è stato invece simulato eliminando gli hub ad uno ad uno, da quello con grado maggiore fino a 50 hub tolti. Come misura della tolleranza ad attacchi ed errori della rete sono state misurate le variazioni di diametro e distanza media in seguito all'eliminazione dei nodi. Sono state inoltre ricavate le rappresentazioni grafiche delle reti ottenute dalla simulazione di attacchi ed errori.

3. ANALISI DELLE INTERAZIONI CON L'OSPITE

Per studiare le interazioni con l'ospite, sono state individuate nella rete le molecole che i due simbionti si scambiano.

Le molecole che il simbiote potrebbe acquisire dall'ospite sono quelle che nella rete metabolica hanno solo link in uscita (cioè figurano solo tra i substrati delle reazioni e non tra i prodotti). Quindi sono stati elencati i nodi che hanno k_{in} pari a 0, come possibili candidati a questo ruolo.

Le molecole che invece vengono fornite dal simbiote all'ospite sono quelle che hanno solo link in entrata nella rete metabolica (cioè sono solo prodotti dalla cellula e non figurano tra i reagenti). Sono stati perciò individuati i nodi con k_{out} pari a 0 come possibili candidati per questo ruolo.

Questa analisi è stata condotta solo sulla rete metabolica di *Wolbachia* simbiote di *Brugia malayi* in quanto essa è laboriosa e richiede molto tempo.

4. COSTRUZIONE DELL'ALBERO FILOGENETICO

Per avere un quadro delle relazioni filogenetiche tra gli organismi in studio è stato costruito un albero filogenetico basato sulle sequenze di 16S rRNA, oltre a sequenze dei geni *groEL* e *ftsZ*.

Per la correzione delle sequenze è stato usato il programma Bioedit 7.0.0 (Hall, 1999). Le sequenze sono state poi allineate con Clustal X (Thompson et al., 1997). Infine per l'analisi filogenetica è stato utilizzato MEGA 3.1 (Kumar et al., 2004). Come approccio all'analisi filogenetica è stato impiegato il metodo /neighbor joining/, un algoritmo che utilizza le distanze esistenti tra ogni paio di sequenze a disposizione per costruire gli alberi. Si è applicata la correzione di Kimura.

5. Risultati e discussione

1. ANALISI DELLA RETE METABOLICA DI *Wolbachia*

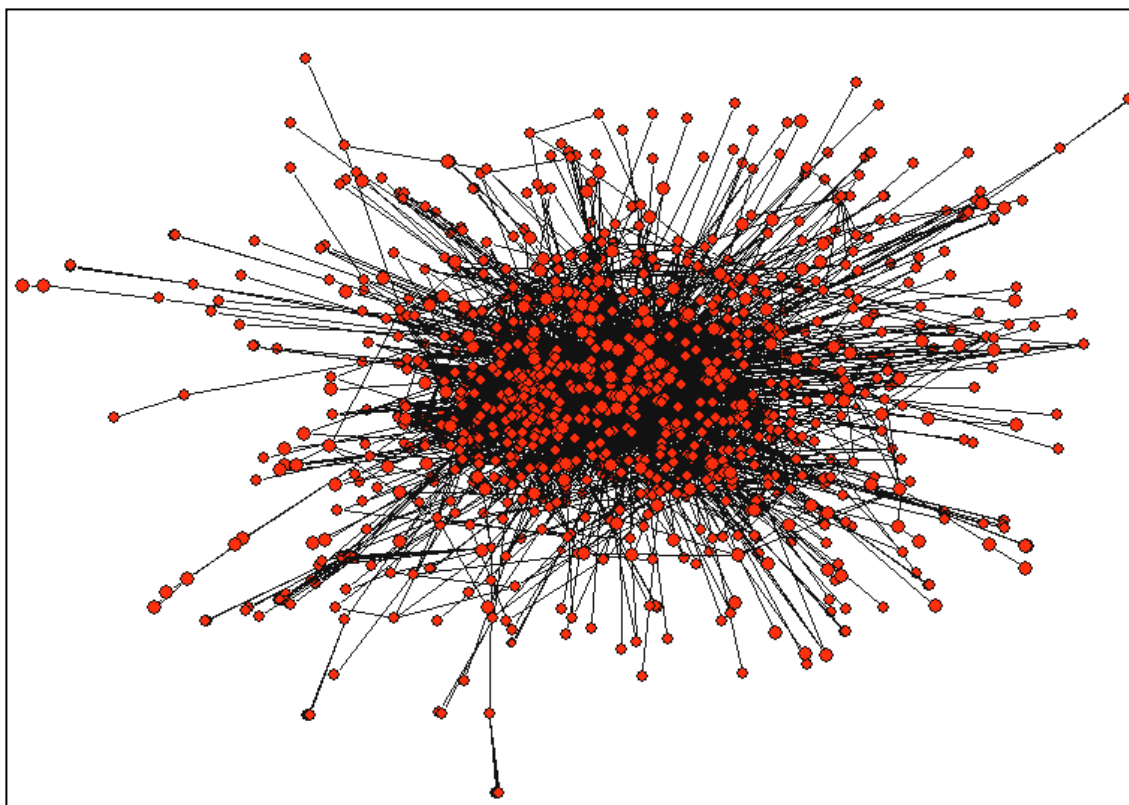


Fig. 9 . La struttura della rete metabolica di *Wolbachia* simbiote del nematode filaride *Brugia malayi*. La rete è stata disegnata utilizzando il programma Pajek (batagelj e Mrvar, 1998).

1.1. Caratteristiche strutturali della rete

Numero di nodi: 2100, di cui 1100 centri di reazione e 1000 substrati.

Numero di link: 8128.

1.1.1. Diametro e cammino medio tra i nodi

Diametro (massima distanza tra ogni coppia di nodi): 16.

Distanza media tra i nodi (gradi di separazione): 5,18331.

Densità (numero di nodi/diametro): 131,25.

Si può notare come, nonostante l'elevato numero di nodi, la rete metabolica di *Wolbachia* abbia la proprietà di piccolo mondo. Infatti in media sono sufficienti 5,18 passaggi per connettere qualsiasi coppia di nodi della rete.

1.1.2. Distribuzione dei gradi dei nodi in entrata e in uscita

Dal momento che la rete metabolica è una rete orientata, per ogni nodo è stato calcolato il numero di link in entrata (k_{in}) e il numero di link in uscita da esso (k_{out}). Per ogni valore di k_{in} e k_{out} è stata quindi calcolata la frequenza e rappresentata poi in grafico (Figura 10).

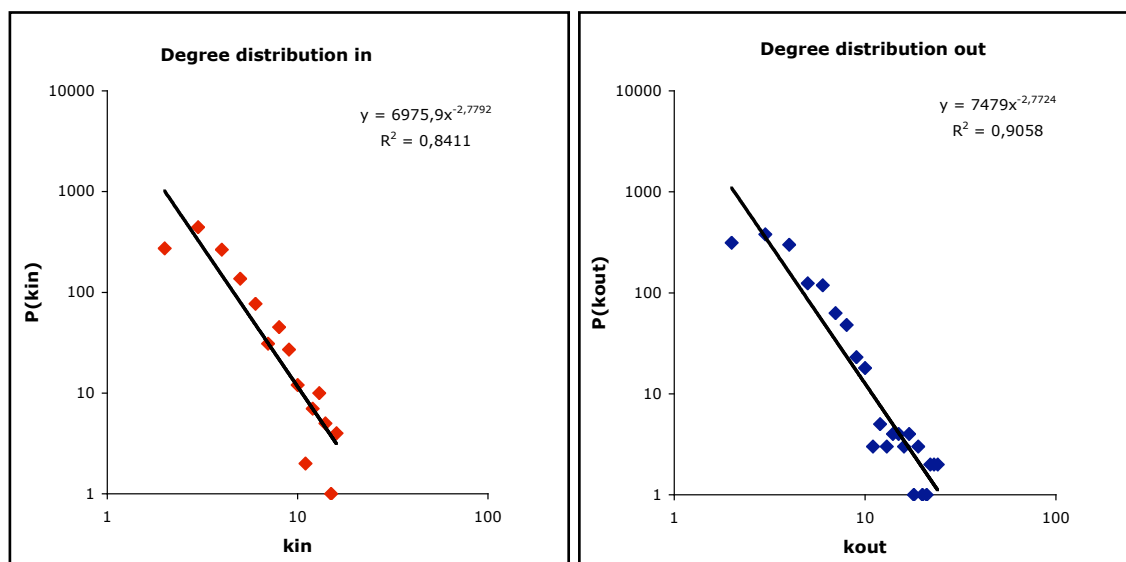


Fig. 10. Distribuzione dei gradi in entrata e in uscita dei nodi della rete metabolica di *Wolbachia*. Gli assi sono in scala logaritmica. Sono mostrate le equazioni della retta di regressione e i coefficienti di regressione.

I grafici mostrano la distribuzione della probabilità che un nodo abbia un certo numero di link in entrata (cioè sia un prodotto delle reazioni cellulari) o in uscita (cioè partecipi come substrato alle reazioni metaboliche). Si può vedere come, in entrambi i casi, ad eccezione che per il primo punto, $k=1$, la probabilità si distribuisce secondo una legge di potenza, cioè la probabilità di trovare un nodo con grado k è proporzionale a $k^{-\gamma}$. Gli esponenti di grado delle due distribuzioni sono $\gamma_{in}= 2,78$ e $\gamma_{out}= 2,77$: ciò significa che ogni volta che il grado del nodo raddoppia, il numero di nodi con questo grado diminuisce di più di 4 volte. Questo significa che la rete metabolica di *Wolbachia* è una rete a invarianza di scala secondo la definizione di Barabasi e Albert (1999).

1.1.3. Hub

La caratteristica principale delle reti a invarianza di scala, conseguenza della distribuzione dei gradi dei nodi secondo una legge di potenza, è la coesistenza nella rete

di molti nodi con pochi legami e un piccolo numero di nodi altamente connessi, gli hub. Questi ultimi sono molto importanti per la connettività della rete stessa. In questo lavoro sono stati dunque identificati i nodi con un numero elevato di link in entrata (hub in) e in uscita (hub out). In particolare, sulla base di quanto osservabile in bibliografia (Wagner e Fell, 2001), si è stabilito che sono hub i nodi aventi un numero di link superiore alla somma tra la media dei valori di k e 3 volte la deviazione standard ($k > \bar{k} + 3\sigma$).

Per quanto riguarda gli hub in, il valore medio di k_{in} è di 3,87 link per nodo. La deviazione standard è pari a 11,49. Perciò sono considerati hub in i nodi aventi un grado k_{in} superiore a $3,87 + (3 * 11,49) = 38,34$. La maggior parte di questi nodi rappresentano molecole il cui ruolo chiave nel metabolismo cellulare è già noto o scontato. Oltre a questi nodi, sono stati perciò considerati hub anche i nodi con un numero di link superiore al limite calcolato in precedenza diminuito di 10, cioè 28. Gli hub in rappresentano i nodi con il maggior numero di link entranti, cioè le molecole che vengono prodotte in un gran numero di reazioni cellulari. Gli hub in individuati sono elencati in tabella 3.

Tab. 3. Elenco dei nodi con un elevato numero di link entranti (hub in).

Hub in	Grado	Hub in	Grado
ADP	365	AMP	64
Ortofosfato	334	S-lattato	53
H ₂ O	266	CoA	43
CO ₂	160	2-oxoglutarato	43
ATP	157	Diidrolipoammide	37
Piruvato	121	NADP+	35
FADH ₂	117	L-glutammato	34
Pirofosfato	104	Glutazione	31
NADPH	80	AcetilCoA	28

Per quanto riguarda gli hub out, il valore medio di k_{out} per la rete metabolica di *Wolbachia* è di 3,87 link per nodo; la deviazione standard è di 9,86. Un hub out ha quindi un grado k_{out} superiore a $3,87 + (3 * 9,86) = 33,45$. Anche in questo caso sono stati consi-

derati anche i nodi con un valore di k_{out} superiore al limite fissato diminuito di 10, cioè 23. Gli hub out rappresentano i nodi con un elevato numero di link in uscita, cioè le molecole che maggiormente partecipano come reagenti nelle reazioni metaboliche cellulari. Gli hub out individuati sono elencati in tabella 4.

Tab. 4. Elenco dei nodi con un elevato numero di link uscenti (hub out).

Hub out	Grado	Hub out	Grado
ATP	404	L-aspartato	37
H ₂ O	372	AcetilCoA	34
ADP	205	Pirofosfato	33
ortofosfato	194	Pcarrier protein	32
D-glucosio	138	Piruvato	29
NAD ⁺	127	Glutazione	28
FAD	47	Coenzima A	26
L-glutamato	40	Glicerolo-3P	24
Alfa-D-glucosio	37	NADPH	24
Lipoammide	37		

Come si può osservare dalle tabelle 3 e 4, in entrambi i casi gli hub sono rappresentati da molecole di cui già era noto il ruolo chiave nel metabolismo, in quanto è risaputa la loro partecipazione alla maggior parte delle reazioni biochimiche cellulari. Rispetto all'analisi degli hub della sola rete di *Wolbachia* è risultato di maggior interesse il confronto tra gli hub di tutte le reti costruite, che verrà presentato nel prossimo capitolo.

1.2. Analisi della tolleranza ad attacchi ed errori

Le natura a invarianza di scala conferisce alle reti un'importante proprietà: la capacità di tollerare errori nella sua architettura. Per molte reti a invarianza di scala è stato infatti dimostrato che eliminando casualmente dei nodi il loro diametro non varia, e la rete non subisce grandi variazioni nella sua struttura. Questo però non accade nel caso di attacchi mirati: se si selezionano e si eliminano gli hub, la rete si disintegra più rapidamente, e le distanze tra i nodi aumentano. Sono state quindi

testate la resistenza e la tolleranza nei confronti di errori casuali ed attacchi mirati della rete metabolica di *Wolbachia*. Per fare ciò sono state misurate le variazioni nel diametro e nella distanza media della rete, dopo aver eliminato dei nodi scelti a caso o gli hub.

Eliminando i nodi in modo casuale non si osserva un aumento significativo dei parametri. Il diametro e la distanza media rimangono all'incirca costanti anche dopo aver tolto il 5% dei nodi (Figura 11). Questo è dovuto al fatto che eliminando casualmente i nodi c'è un'alta probabilità di colpire nodi con pochi link, che quindi non hanno un ruolo essenziale nel definire e mantenere la connettività della rete. Una particolarità è data dalla diminuzione del diametro in seguito all'eliminazione dei nodi: una possibile interpretazione di questo risultato potrebbe essere l'eliminazione di nodi periferici, con il conseguente isolamento di gruppi. Il diametro quindi verrebbe misurato solo sul gruppo centrale di nodi, che risulterà più piccolo per l'eliminazione dei gruppi più periferici.

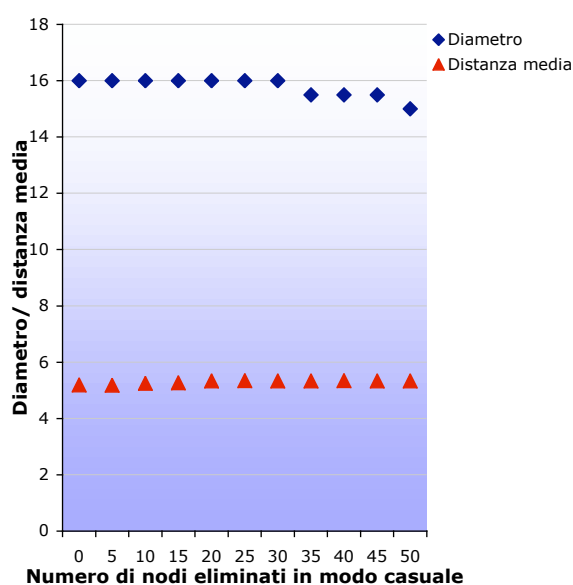


Fig. 11. Variazione del diametro e della distanza media in seguito all'eliminazione di N nodi scelti in modo casuale. In totale sono stati eliminati 50 nodi, rappresentanti il 5% dei substrati della rete.

Eliminando gli hub uno alla volta, da quello con grado maggiore, si è invece osservato un aumento sia del diametro che della distanza media tra i nodi (Figura 12). Dopo aver eliminato 10 hub (che rappresentano l'1% dei nodi della rete) si ha un aumento notevole del diametro: la distanza massima tra due nodi passa da 16 a 22 passaggi.

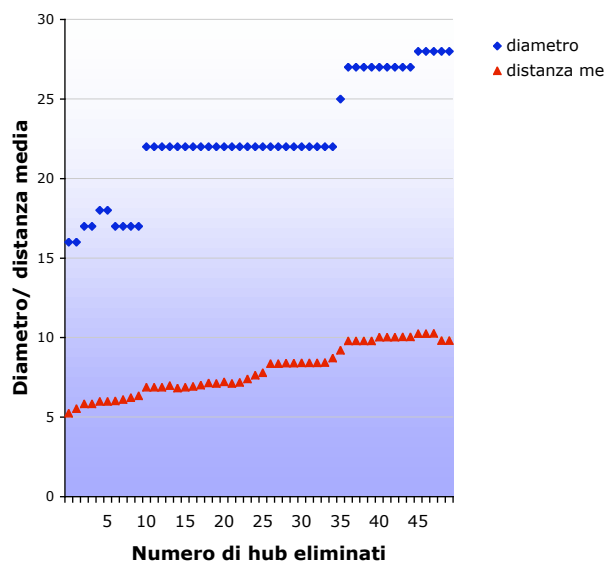


Fig. 12. Effetto dell'eliminazione degli hub sul diametro e la distanza media. In totale sono stati eliminati 50 hub.

L'aumento del diametro e della distanza media in seguito all'eliminazione degli hub è significativo dell'importanza di questi nodi per la connettività della rete. Dal momento che un hub è connesso a moltissimi altri nodi, questo permette anche a nodi isolati di ridurre la loro distanza dagli altri nodi connessi all'hub.

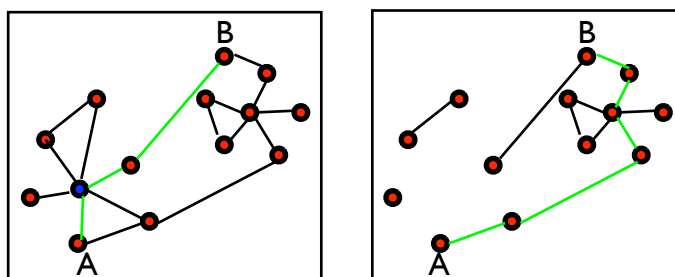
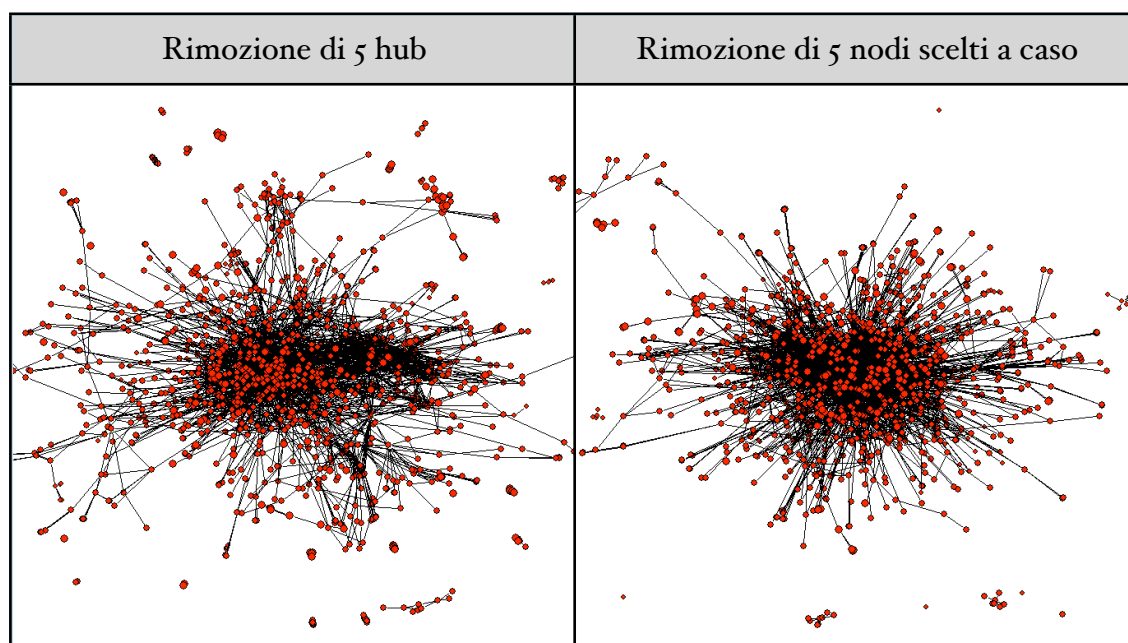
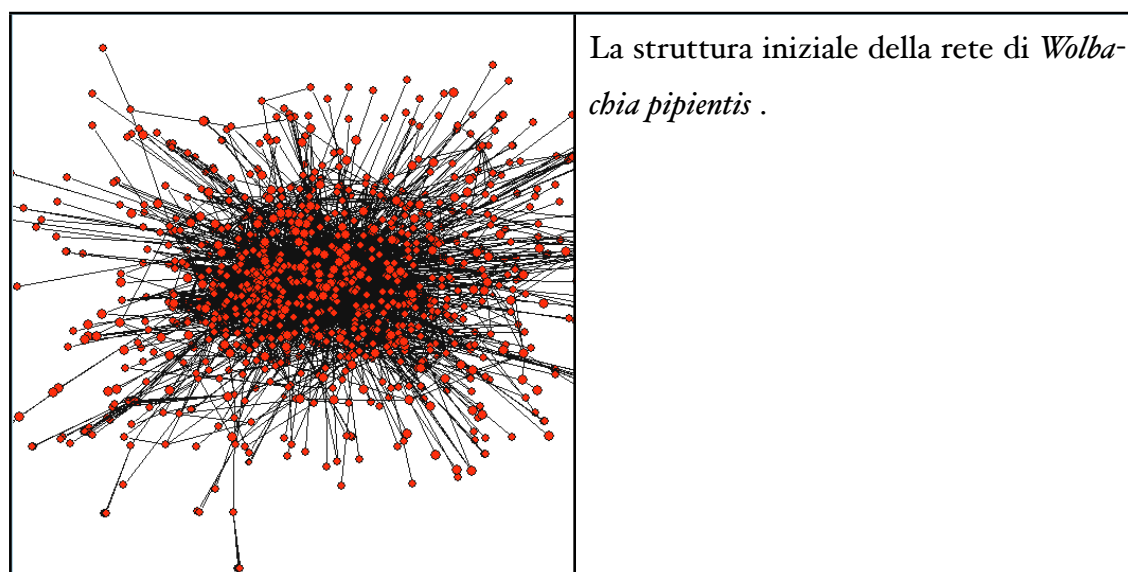


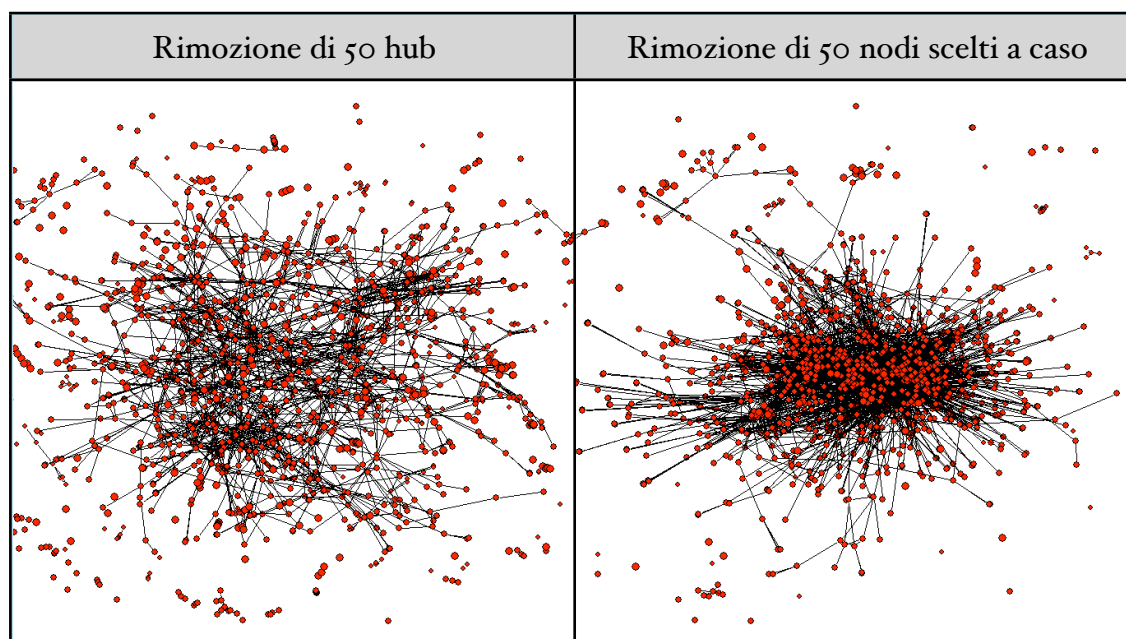
Fig. 13. Effetto dell'eliminazione di un hub sulla connettività della rete.

Come si vede in figura 13, eliminando l'hub (nodo blu) il cammino tra i nodi A e B cresce da 3 passaggi a 5. La rete metabolica di *Wolbachia* ha quindi, come tutte le reti a invarianza di scala, una ridotta capacità di resistenza agli attacchi mirati contro gli hub.

Anche dall'osservazione della struttura delle reti si può notare la differenza degli effetti ottenuti eliminando gli hub o un numero di nodi scelti a caso.



La struttura della rete di *Wolbachia pipientis* dopo aver eliminato i 5 hub con grado maggiore e 5 nodi scelti a caso. L'attacco selettivo agli hub ha modificato l'architettura della rete, generando dei vuoti visibili nell'intrico dei link. Inoltre si forma già un certo numero di isole, gruppi di nodi non più collegati al gruppo centrale. Con l'eliminazione di nodi casuali non si hanno invece grosse conseguenze sulla topologia della rete.



Dopo aver eliminato 50 nodi scelti a caso (il 5% sul totale) aumentano i nodi isolati, mentre il centro della rete subisce minori conseguenze. Eliminando invece i 50 nodi più connessi la rete si frammenta in molti piccoli gruppi di nodi isolati e nel centro della rete la densità dei link si riduce visibilmente.

Si è visto che gli hub sono rappresentati da molecole che partecipano alla maggior parte delle reazioni biochimiche che avvengono nella cellula e che quindi sono importanti per la sopravvivenza della cellula stessa. Questo è stato confermato dall'analisi della tolleranza ad attacchi ed errori della rete. La topologia a invarianza di scala, infatti, conferisce alla rete la capacità di reagire positivamente ad errori che coinvolgono nodi poco connessi, quindi molecole che partecipano a poche reazioni. D'altra parte, però, se si colpissero gli hub in primo luogo aumenterebbe il diametro, quindi la cellula impiegherebbe più tempo ad adattarsi a un cambiamento ambientale. Inoltre l'architettura stessa della rete ne risentirebbe, al punto di frammentarsi in tanti gruppi di nodi isolati. Il ruolo chiave delle molecole-hub è quindi anche quello strutturale di mantenere unita l'architettura della rete.

1.3. Analisi delle interazioni con l'ospite

L'obiettivo di questo tipo di analisi è quello di identificare molecole che potrebbero venire scambiate tra ospite e simbionte. Il risultato sono due liste di molecole: quelle rappresentate da nodi con solo link in uscita, cioè che partecipano solo come reagenti nel metabolismo, e quelle rappresentate da nodi con solo link in entrata, che sono solo prodotte dalla cellula. Nelle tabelle sono elencate le molecole rilevate dal-

l'analisi della rete metabolica di *Wolbachia* e raggruppate in base alla classe chimica o al pathway metabolico di appartenenza. Si suppone che le molecole che figurano solo tra i reagenti della rete metabolica di *Wolbachia* vengano prelevate dall'ambiente esterno, cioè dalle cellule dell'ospite. Le molecole che compaiono solo tra i prodotti, invece, potrebbero essere, oltre a sostanza di rifiuto, anche composti che il batterio fornisce all'ospite. I dati ottenuti sono quindi stati confrontati con le informazioni sugli scambi tra simbionte ed ospite che già si hanno, derivate dall'analisi dei loro genomi.

Tab. 5. Molecole che figurano solo tra i reagenti, cioè che non hanno link in entrata, nella rete metabolica di *Wolbachia pipientis*.

Possibili molecole acquisite dall'ospite	
Vitamine/cofattori	Riboflavina (B2), Mio-inositolo (B), quinate, ferredoxina
Glicolisi /TCA	beta-D-glucosio; alfa-D-glucosio-1P; UDP glucosio, D-fruttosio, ribosio, D-ribosio-1P, 2-D-Pglicerato, Gliceraldeide, 2-D-Pglicerato, D-gluconato, 2-deidro-D-gluconato; citrato, isocitrato, defosfoCoA
Aminoacidi	aspartato-semialdeide, L-alanina, L-arginina, L-asparagina, L-cisteina, glicina, L-istidina, L-isoleucina, L-leucina, Tirosina, L-lisina, L-valina, L-treonina, L-metionina, Metiltetraidropteroiltri-L- glutammato, L-ornitina, L-fenilalanina, Carbossietil-L-arginina, Idrossi-L-prolina, L-serina, L-allotreonina, D-glutammato, D-alanina

Tab. 6. Molecole che figurano solo tra i prodotti, cioè che non hanno link in uscita, nella rete metabolica di *Wolbachia pipientis*.

Possibili molecole fornite all'ospite	
Vitamine/cofattori	Mio-inositolo (gruppo B), Diidrofolato (B9), piridoammina (B6), Adenosilcobalammina (B12), Menachinone (K) Tetraidrometanopterinina, pirrolochinoline-chinone, molibdotterina, F420(red), FADH, diidrometanopterinina, ferrocitocromo-b, Protome, sirome
Glicolisi /TCA	Alfa-D-glucosio-1,6biP, Alfa-D-ribosio-1P, Glutarato, Malato Lattato, Etanolo, Butanolo
Metabolismo degli acidi grassi	Acetone, propionil-CoA, palmitate, palmitoil-CoA, butanoil-CoA
Aminoacidi	Glucosammina, tiocisteina, omocisteina, ornitina
Nucleotidi	AMP, UMP, UDP, TDP, CMP, CDP, NTP, GDP, ITP, dITP, adenina, guanina, inosina, ipoxantina, D-alanilil-D-alanina, adenosina, nicotinamide_D-ribonucleotide, 4-(1-D-ribitylamino)-5-amino-2,6-dihidropirimidine, nicotinamide_nucleotide, 5'-extraoligonucleotide, 3'-phosphomononucleotide

Dai dati disponibili finora riguardo al rapporto tra *Wolbachia* e il nematode *Brugia malayi* si pensa che il batterio acquisisca dall'ospite cofattori e vitamine (CoA, NAD, biotina, ubiquinone e altri), oltre a substrati di crescita (zuccheri e amminoacidi). A sua volta, l'ospite dipenderebbe da *Wolbachia* per i nucleotidi, e alcuni coenzimi (riboflavina, FAD), oltre che per l'eme. I risultati dell'analisi della rete metabolica confermano questi dati: infatti molti precursori metabolici e substrati della glicolisi sono stati rinvenuti nella lista dei nodi con solo link in uscita (reagenti), oltre ad amminoacidi: dato che *Wolbachia* ha come ambiente di crescita la cellula dell'ospite, acquisirà da questo i substrati di crescita, oltre ad alcuni cofattori. Per quanto riguarda le molecole trasferite all'ospite, i risultati confermano le ipotesi ricavate dall'analisi del genoma: il batterio produce nucleotidi, vitamine e cofattori, tra cui l'eme, oltre ad altre molecole, tra le quali ovviamente si riscontrano prodotti metabolici di rifiuto.

L'analisi presentata evidenzia quindi come il metodo di analisi proposto, permetta di identificare delle note molecole implicate nei processi simbiotici tra gli organismi in esame. Parallelamente vengono identificate anche altre molecole che per il momento non sono note come implicate in processi della simbiosi. Riteniamo però che le altre molecole identificate in questo modo potrebbero essere possibili candidati su cui basare studi futuri. In questa sede si è voluto proporre un nuovo metodo per lo studio delle relazioni ospite-simbionte, basato sull'analisi delle loro reti metaboliche. Una conferma della correttezza di questa analisi potrebbe derivare dal confronto tra le reti metaboliche dei due simbionti. Se si trovassero corrispondenze tra le molecole che figurano solo tra i reagenti del simbionte e solo tra i prodotti dell'ospite, e viceversa, si potrebbe avere conferma del fatto che queste sostanze effettivamente vengono scambiate tra i due organismi. Al momento non sono però disponibili dati riguardo ai pathway metabolici di *Brugia malayi*.

2. ANALISI E CONFRONTO DELLE RETI METABOLICHE

2.1. Relazioni filogenetiche

In totale sono state ricostruite ed analizzate le reti metaboliche di 18 microrganismi, di cui 14 endosimbionti (mutualisti e parassiti) e 4 a vita libera. Allo scopo di chiarire le relazioni filogenetiche tra gli organismi in esame è stato costruito un albero filogenetico basato sulle sequenze di rRNA 16S.

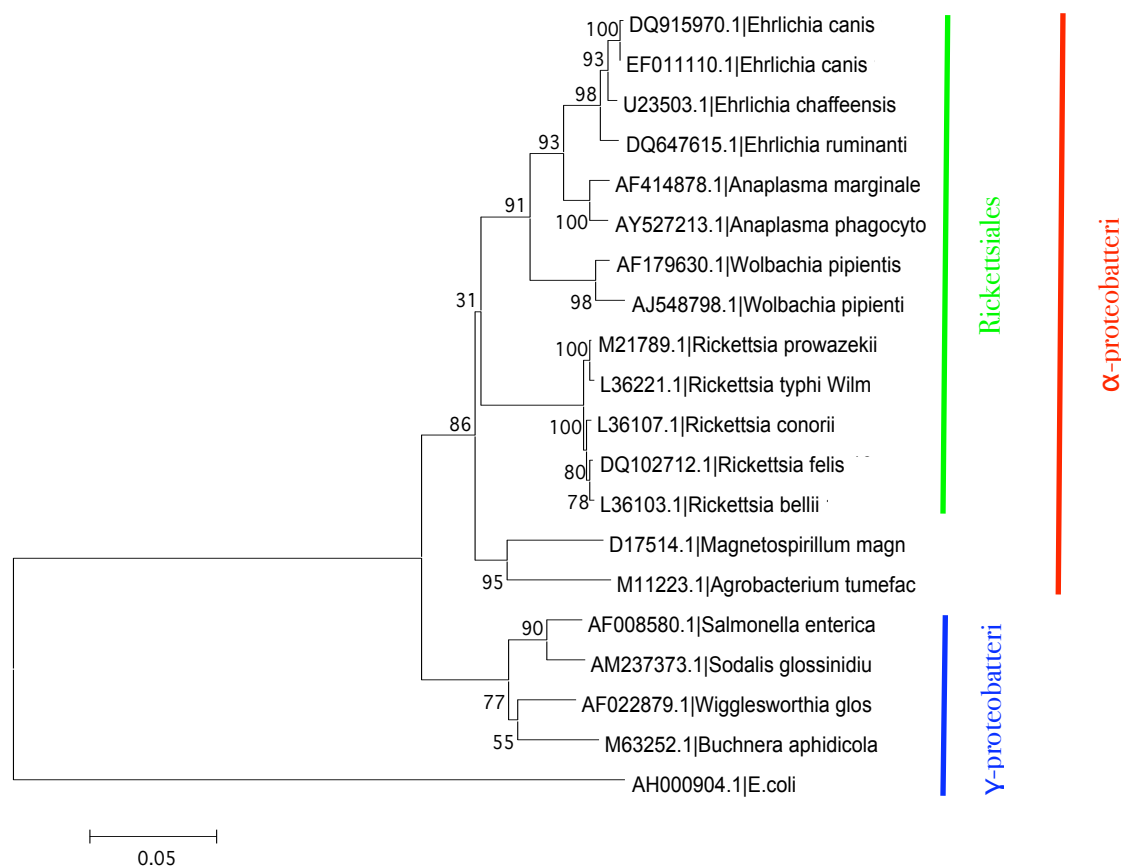


Fig. 14. Albero filogenetico (neighbour joining) basato sulle sequenze di rRNA 16S.

L'albero ottenuto ha confermato la suddivisione dei batteri oggetto di studio nei due gruppi di proteobatteri, α e γ . Inoltre, i generi di batteri simbiotici *Ehrlichia*, *Anaplasma*, *Wolbachia* e *Rickettsia* costituiscono un monophylum, in quanto parte dell'ordine dei Rickettsiales. In questo gruppo le specie di *Rickettsia* risultano separate rispetto agli altri tre generi. Le ricostruzioni basate sulle sequenze disponibili dei geni groEL e ftsZ non presentano differenze significative rispetto a quella presentata.

2.2. Analisi delle reti metaboliche

Per ogni rete è stato condotto lo stesso tipo di analisi utilizzato per la rete di *Wolbachia pipientis*, comprendente la misura del diametro e della distanza media tra i nodi, lo studio della distribuzione dei gradi dei nodi, sia in entrata che in uscita, e l'identificazione degli hub. In tabella 7 sono riassunte le caratteristiche strutturali delle reti ottenute.

Tab. 7. Caratteristiche strutturali delle reti metaboliche costruite per i 18 organismi. Sono indicati anche lo stile di vita dei batteri e la dimensione del loro genoma. V= stile di vita, S= substrati, CR= centri di reazione, D= diametro, DM= distanza media, P= parassita, M= mutualista, L= vita libera.

	Organismo	V	Genoma	Nodi	S	CR	Link	γ_{in}	γ_{out}	D	DM
1	<i>A. phagocytophilum</i>	P	1471 kb	2056	1015	1041	7924	2,69	3,60	14	5,08
2	<i>A. marginale</i>	P	1197 kb	2316	1115	1201	8913	2,16	2,24	14	5,18
3	<i>R. typhi</i>	P	1111 kb	1874	986	888	6270	2,68	2,70	16	5,44
4	<i>R. prowazekii</i>	P	1111 kb	1898	1023	875	6563	2,59	2,73	14	5,28
5	<i>R. felis</i>	P	1485 kb	2062	1086	976	6966	2,74	2,75	14	5,30
6	<i>R. conorii</i>	P	1268 kb	1881	979	902	6603	2,71	2,75	14	5,28
7	<i>R. bellii</i>	P	1522 kb	1977	1052	925	6704	2,05	2,74	14	5,30
8	<i>E. canis</i>	P	1315 kb	2265	1087	1178	8702	2,80	2,50	14	5,17
9	<i>Wolbachia</i>	M	1080 kb	2100	1000	1100	8128	2,78	2,77	16	5,18
10	<i>W. glossimidia</i>	M	697 kb	2277	1103	1174	8399	2,73	2,84	14	5,21
11	<i>S. glossinidius</i>	M	4172 kb	3185	1466	1719	11735	2,67	2,49	15	5,29
12	<i>B. aphidicola Sg</i>	M	641 kb	2342	1140	1202	8815	2,73	2,96	14	5,18
13	<i>B. aphidicola Bp</i>	M	615 kb	2251	1094	1157	8456	2,74	2,08	14	5,15
14	<i>B. aphidicola Ap</i>	M	640 kb	2371	1152	1219	9003	2,75	2,99	14	5,18
15	<i>S. enterica</i>	L	4755 kb	3717	1686	2031	13309	2,62	2,85	15	5,34
16	<i>M. magneticum</i>	L	4967 kb	3052	1438	1614	11306	2,75	2,85	14	5,26
17	<i>E. coli</i>	L	4938 kb	3844	1756	2088	13598	2,73	2,80	15	5,36
18	<i>A. tumefaciens</i>	L	4915 kb	3505	1600	1905	12862	2,15	2,92	15	5,28

2.1.1. Dimensione delle reti metaboliche

Le reti metaboliche dei diversi organismi sono state confrontate per le loro dimensioni, cioè per il numero dei nodi (Figura 15). In grafico sono riportate le dimensioni delle reti ricostruite, suddivise in base allo stile di vita degli organismi.

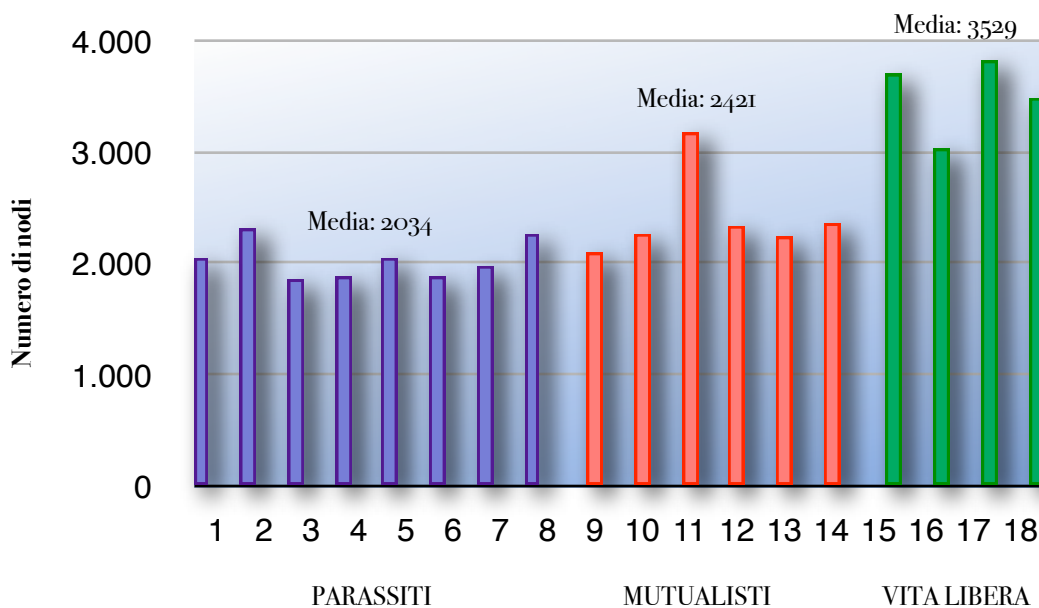


Fig. 15. Dimensione delle reti ricostruite, suddivise sulla base dello stile di vita degli organismi. I numeri in ascissa indicano gli organismi così come elencati in tabella 7.

Dal grafico emerge che le reti metaboliche dei simbiotici hanno un numero di nodi inferiore rispetto a quelle dei batteri a vita libera, con valori medi per i due gruppi che differiscono di circa 1300 nodi (media del numero dei nodi per i simbiotici: 2228). Un'eccezione è rappresentata da *Sodalis glossinidius*, simbiote mutualista (n.11 nel grafico) la cui rete è formata da 3185 nodi, un valore più vicino a quello delle reti dei batteri a vita libera, rispetto a quello dei simbiotici. Non si riscontrano, invece, differenze significative nella dimensione delle reti tra organismi mutualisti e parassiti.

Questi dati sono stati confrontati con quelli riguardanti la dimensione dei genomi (Figura 16). Anche in questo grafico sono stati separati gli organismi in base allo stile di vita.

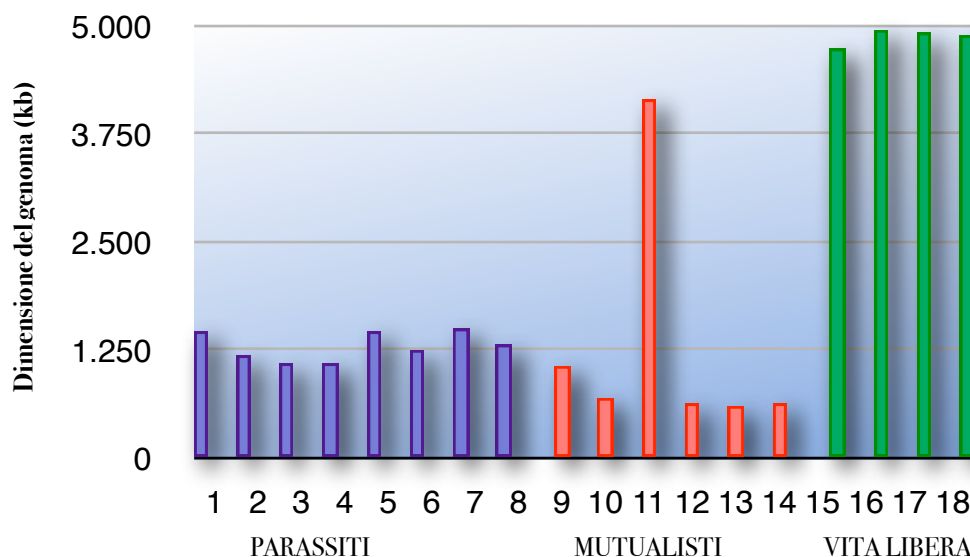


Fig 16. Dimensioni dei genomi degli organismi in esame. I numeri in ascissa indicano gli organismi, come da tab. 7.

Confrontando i due grafici, si può notare una corrispondenza tra la dimensione del genoma e quella della rete: gli organismi a vita libera hanno un genoma più grande rispetto ai simbionti, e questo porta ad una rete di dimensioni maggiori. Ai simbionti, invece, con un genoma di dimensioni minori, risultato del processo di riduzione genomica, corrisponde una rete più piccola, ad eccezione di *S. glossinidius* (n.11). Particolarmente interessante, però, è il fatto che tra batteri a vita libera e simbionti vi è una maggiore differenza nella dimensione dei genomi che nella dimensione della rete. I genomi dei batteri a vita libera sono circa 4 volte più grandi di quelli dei simbionti, mentre il numero dei nodi è meno del doppio nei non simbionti. Per quantificare questa informazione, è stato calcolato, per ogni organismo, il rapporto tra la dimensione della rete metabolica e quella del genoma (Tabella 8).

Tab. 8. Rapporto tra numero di nodi della rete metabolica e dimensione del genoma (in kb).

Parassiti		Mutualisti		A vita libera	
<i>A.phagocytophilum</i>	1,4	<i>W.pipientis</i>	1,94	<i>S.enterica</i>	0,78
<i>A.marginale</i>	1,93	<i>W.glossinidia</i>	3,27	<i>M.magneticum</i>	0,61
<i>R.typhi</i>	1,69	<i>S.glossinidius</i>	0,76	<i>E.coli</i>	0,78
<i>R.prowazekii</i>	1,66	<i>B.aphidicola Sg</i>	3,65	<i>A.tumefaciens</i>	0,71
<i>R.felis</i>	1,39	<i>B.aphidicola Bp</i>	3,66		
<i>R.conorii</i>	1,48	<i>B.aphidicola Ap</i>	3,7		
<i>R.bellii</i>	1,3				
<i>E.canis</i>	1,72				

Da questi dati si può osservare come il rapporto tra la dimensione del genoma e il numero di nodi della rete sia inferiore a 1 per tutti gli organismi a vita libera, ma anche per *S. glossinidius*. La situazione di questo batterio è particolare. Infatti, nonostante la natura da endosimbionte, il suo genoma è rimasto di dimensioni elevate, comparabili a quelle di batteri a vita libera, e lo stesso vale per la dimensione della rete metabolica. Questo potrebbe essere un segnale del fatto che la simbiosi di *S. glossinidius* con la mosca tsetse si è instaurata in tempi recenti. Nei parassiti il rapporto cresce, ma si mantiene al di sotto di 2. Infine, nei mutualisti si osservano le maggiori differenze: mentre *S. glossinidius* è assimilabile ad un batterio a vita libera anche per il valore di questo parametro, i “veri” mutualisti, cioè i tre ceppi di *B. aphidicola* e *W. glossinidia* hanno un rapporto numero dei nodi/genoma superiore a 3, a indicare un genoma molto piccolo che produce però una rete con un elevato numero di nodi, in rapporto. *Wolbachia*, invece, ha un rapporto inferiore a 2, dunque più vicino a quelli misurati per i parassiti: in effetti il suo genoma ha subito una minore riduzione in confronto a quelli di *B. aphidicola* e *W. glossinidia*. I batteri mutualisti sono, tra gli organismi analizzati, quelli che hanno subito un maggiore processo di *genome shrinkage*, con dimensioni del genoma inferiori alle 700 kb. Anche i parassiti hanno subito una riduzione, ma hanno comunque mantenuto un genoma più grande rispetto ai mutualisti, con dimensioni superiori alle 1000 kb. Nonostante questa differenza di dimensioni del genoma, si può però osservare come nelle due classi di simbiotici non vi siano differenze riguardo alla dimensione della rete metabolica. Questo potrebbe significare che, nonostante la riduzione del genoma seguita all’instaurarsi della simbiosi, vi sia stata una tendenza a conservare i geni coinvolti nelle reazioni metaboliche cellulari.

2.2.2. Diametro e cammino medio tra i nodi

Nella tabella 7 sono elencati i valori del diametro e della distanza media tra i nodi misurati per le reti metaboliche di tutti gli organismi in esame. Questi valori sono molto simili per le reti di tutti gli organismi analizzati: non vi sono differenze legate al tipo di vita dei batteri o alle loro relazioni filogenetiche. Nel grafico (Figura 17) sono rappresentati i valori del diametro misurati, in funzione della dimensione della rete: si può osservare come non ci sia una variazione significativa di questo parametro al variare del numero dei nodi della rete. Per diametro si intende la distanza massima misurata tra tutte le coppie di nodi della rete.

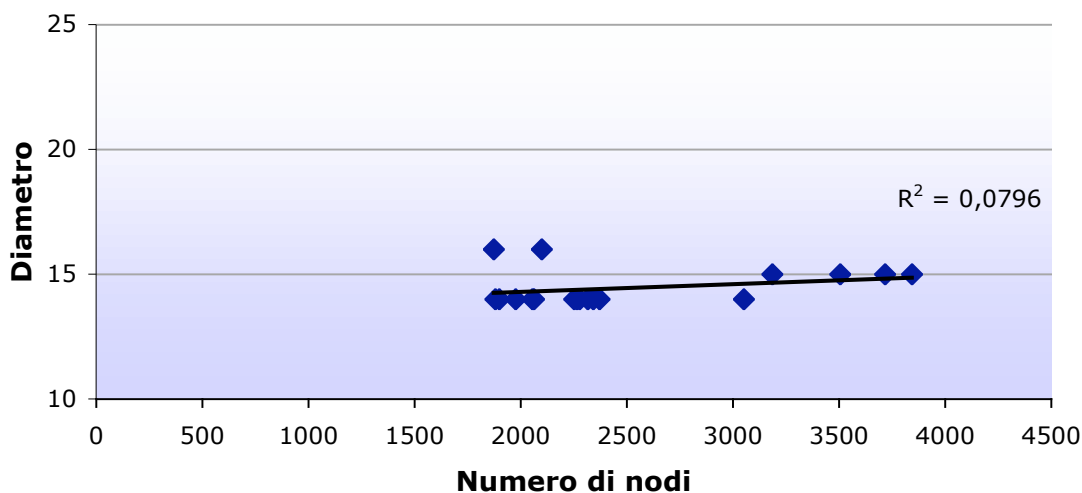


Fig. 17. Variazione del diametro della rete in funzione del numero dei nodi.

In una rete metabolica la distanza tra due nodi rappresenta il numero di passaggi che occorrono per ottenere, da un dato substrato, un certo prodotto. Essendoci diversi possibili pathways di conversione tra due metaboliti, per l'analisi delle reti è importante considerare il più breve tra questi pathways. Anche la distanza media misurata per le reti in studio è indicata in tabella 7. Il valore misurato è molto basso, variando tra 5,1-5,5 per tutte le reti: le reti metaboliche dei batteri presi in esame sono dunque reti piccolo mondo, con un basso grado di separazione. Anche in questo caso non ci sono differenze in questo parametro in relazione alla dimensione della rete, come mostrato dal grafico (Figura 18).

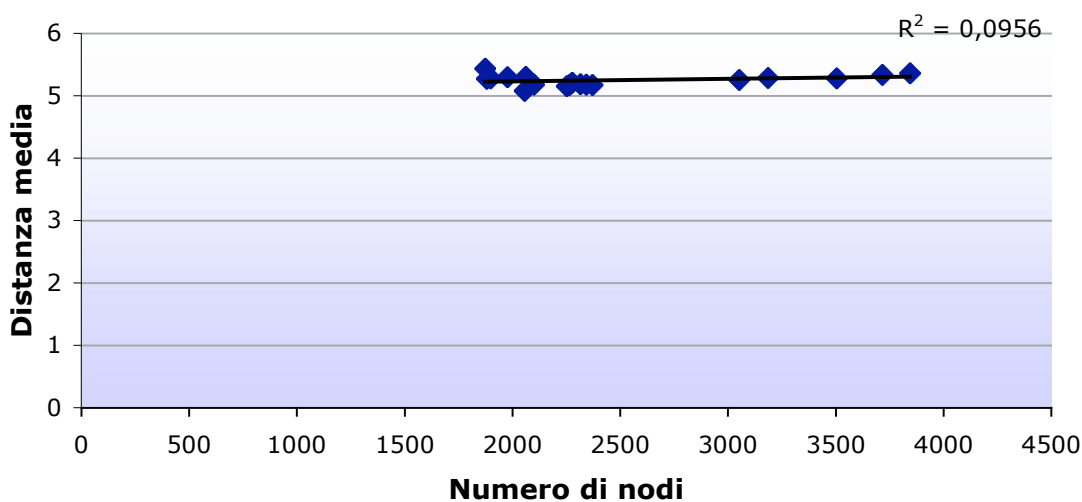


Fig. 18. Variazione della distanza media tra i nodi in funzione della dimensione della rete.

Questo dato è in contrasto con la teoria, per cui il diametro di una rete cresce con il logaritmo del numero dei nodi, verificata in diverse reti non biologiche. L'assenza di variazione del diametro e del cammino medio tra i nodi in funzione della dimensione della rete è stata però rilevata anche nel lavoro di Jeong et al. (2000), dove sono state analizzate le reti metaboliche di 43 organismi appartenenti a domini diversi. Ma e Zeng (2003), invece, hanno misurato un aumento del diametro con le dimensioni della rete. Anche gli organismi presi in esame nel loro studio appartengono sia ad eucarioti che ad archebatteri e batteri. Nella presente indagine, invece, sono stati presi in esame solo organismi procarioti, anche molto vicini filogeneticamente: questo potrebbe essere un motivo che spiegherebbe l'uniformità del diametro delle reti metaboliche analizzate.

Questo aspetto delle reti metaboliche è molto importante: un basso grado di separazione tra le molecole fa sì che una perturbazione in un punto della rete venga trasmessa rapidamente a tutti gli altri nodi e non resti invece localizzata. In questo modo, in seguito ad un cambiamento ambientale, l'informazione passa rapidamente a tutti i nodi della rete, permettendo quindi alla cellula di adattarsi rapidamente.

Dato che il diametro della rete resta costante, all'aumentare della dimensione della rete, i nodi saranno maggiormente addensati: la densità, calcolata come rapporto tra il numero dei nodi e il diametro, aumenta infatti linearmente con il numero dei nodi della rete, come si può osservare dal grafico (Figura 19).

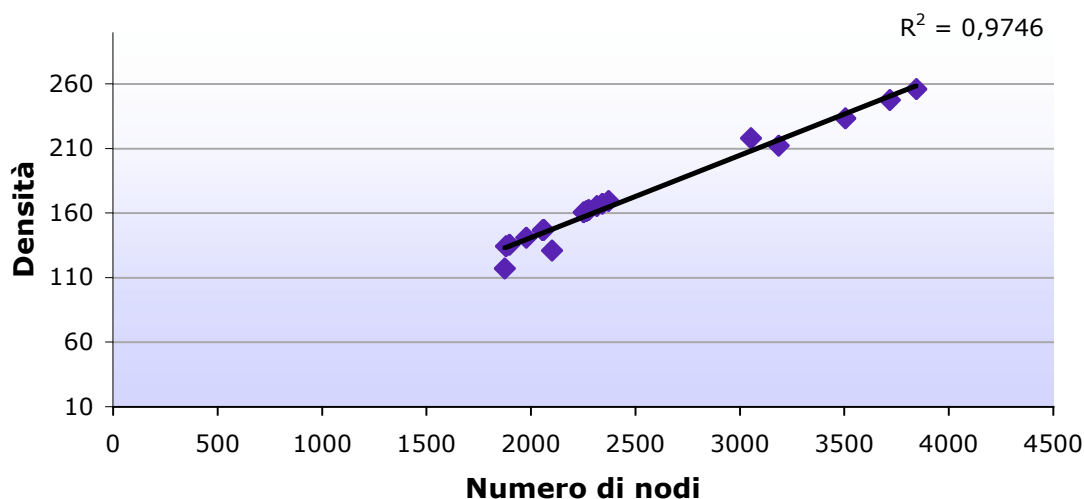


Fig. 19. Variazione della densità (numero dei nodi/ diametro) in funzione della dimensione della rete.

2.2.3. La distribuzione dei gradi dei nodi

La distribuzione dei gradi dei nodi delle reti metaboliche costruite è stata rilevata con la stessa procedura utilizzata per la rete di *Wolbachia*, considerando separatamente i link in entrata e quelli in uscita da ogni nodo della rete. I grafici delle distribuzioni dei gradi dei nodi in entrata e in uscita per tutti gli organismi sono presentati in appendice. Dal momento che non si riscontrano differenze nelle distribuzioni di probabilità, vengono qui presentati i grafici relativi alle reti di tre organismi, rappresentativi delle tre modalità di vita analizzate: un parassita (*Anaplasma marginale*), un mutualista (*Wigglesworthia glossinidia*) e un batterio a vita libera (*Salmonella enterica*).

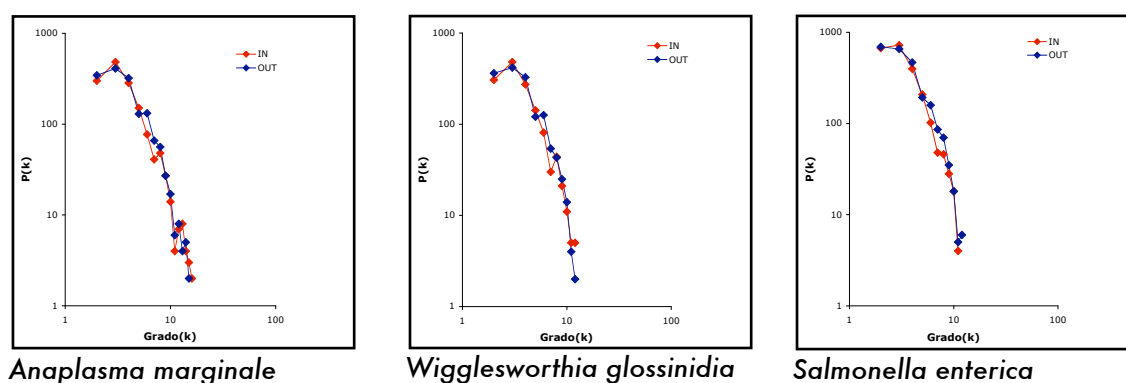


Fig. 20. Distribuzione dei gradi dei nodi (in e out) nelle reti metaboliche di tre organismi studiati, rappresentativi dei tre stili di vita analizzati. Gli assi sono in scala logaritmica.

Dai grafici emerge che in tutte le reti, la distribuzione dei gradi dei nodi, sia per i link in entrata che per quelli in uscita, segue una legge di potenza, con esponenti di grado che variano tra 2 e 3, come riassunto in tabella 7. Tutte le reti metaboliche ricostruite, quindi, sono delle reti a invarianza di scala.

2.2.4. Hub

Seguendo la stessa procedura usata per la rete di *Wolbachia*, anche per le altre reti sono stati individuati gli hub in e gli hub out. L'elenco degli hub per tutte le reti è presentato in appendice. Alcune molecole si sono rivelate essere hub per tutte le reti analizzate: esse sono elencate in tabella 9.

Tab. 9. Elenco delle molecole che figurano tra gli hub di tutte le reti metaboliche.

Molecole hub in per tutte le reti	Molecole hub out per tutte le reti
ADP, H ₂ O, Ortofosfato, Pirofosfato, ATP, CO ₂ , Piruvato, FADH ₂ , 2-oxoglutarato, diidrolipoammide	ATP, H ₂ O, ADP, NAD ⁺ , NADPH, FAD, L-glutamato, Piruvato

Nelle tabelle 10 e 11 sono elencate le molecole identificate come hub in e hub out che presentano differenze nelle reti metaboliche studiate. Il “+” sta ad indicare che la molecola è un hub per la rete dell’organismo, il “-” indica che la molecola non è stata identificata come hub per quell’organismo.

Tab. 10. Elenco degli hub in delle reti analizzate. Gli organismi sono indicati con il numero, come da tabella 7.

Hub In	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
NADH	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+
S-lattato	+	+	-	-	+	-	-	+	+	+	+	+	+	+	+	+	+	+
CoA	+	+	+	+	+	+	+	+	+	-	+	+	+	+	-	+	+	+
NADP ⁺	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+
AMP	-	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+
Glutatione	+	+	-	-	-	-	-	-	+	+	-	+	+	+	-	+	-	-
NH ₃	-	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+
succinato	+	-	+	+	+	+	+	-	-	-	+	-	-	-	+	+	-	+
NADPH	-	-	+	-	-	+	+	-	+	-	+	-	-	-	+	+	+	+
NAD ⁺	-	+	-	-	-	-	-	+	-	+	+	+	-	+	+	+	-	+
L-glutamato	+	+	-	-	+	+	+	-	+	+	+	+	+	+	+	+	+	+
D-gliceraldeide-3P	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	-	+	+
Acetato	-	-	+	-	+	-	+	-	-	-	+	+	+	-	+	+	+	+
AcetilCoA	+	-	+	+	+	-	+	+	-	-	+	-	-	-	+	+	-	+

Tab. 11. Elenco degli hub out per le reti analizzate. Gli organismi sono elencati con il numero, come da tabella 7.

Hub Out	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Ortofosfato	+	+	+	-	+	-	+	+	+	+	+	+	+	+	+	+	+	+
D-glucosio	+	+	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+
lipoammide	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+
acetilCoA	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	-
Glutazione	+	+	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+
Glicina	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+	-	+
CO ₂	-	-	-	+	-	-	-	-	-	-	+	-	-	-	+	+	+	+
Oxalacetato	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
Glicerolo-3P	-	+	+	-	+	-	+	+	+	+	-	-	-	-	-	+	-	-
L-aspartato	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	-
2-oxoglutarato	-	-	+	-	+	+	+	+	-	-	+	+	-	-	-	+	+	+
chorismate	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+
L-glutammina	-	+	-	-	-	-	-	-	-	+	+	-	+	+	+	+	+	+
P-carrier protein	+	+	-	-	-	-	-	+	+	+	-	+	+	+	+	+	+	+
5 α Dribo- siordiP	-	-	-	-	-	-	-	-	-	-	+	+	-	+	+	+	+	+
L-serina	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-
NADP ⁺	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-
CoA	+	+	+	+	+	+	+	+	+	-	+	-	-	-	+	+	+	+
NADH	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
O ₂	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	+	+
L-arginina	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	+	+

I principali hub delle reti metaboliche (tabella 9) sono rappresentati dalle stesse molecole in tutti gli organismi, indipendentemente dal loro stile di vita. Si tratta di sostanze che prendono parte alla maggior parte delle reazioni cellulari, come ATP, ADP, fosfato e cofattori, oltre a precursori metabolici, piruvato, 2-oxoglutarato e glutammato. Il ruolo chiave di questi composti nelle reazioni cellulari era già riconosciuto. Questi sono stati individuati come hub anche in studi precedenti sulle reti metaboliche (Jeong et al., 2000; Wagner e Fell, 2001; Ma e Zeng, 2003). Come si è

visto per la rete di *Wolbachia*, queste molecole hanno anche un ruolo strutturale, in quanto responsabili della connettività della rete e del suo funzionamento.

Osservando gli hub non comuni a tutte le reti (tabelle 10 e 11), si osservano maggiori differenze. Queste possono essere dovute innanzitutto a una differenza nella dimensione della rete, che porta a far variare i parametri calcolati per definire un hub, media e deviazione standard del grado dei nodi. Si può spiegare con questa ragione il maggior numero di hub delle reti di batteri a vita libera (dal n. 15 al 18 in tabella), che possiedono un numero più elevato di nodi. Lo stesso vale per *S. glossinidius* (n.11), mutualista ma con una rete più simile a quelle dei batteri non simbiotici.

Se si osservano, però, gli hub di reti di dimensioni confrontabili, come quelle dei parassiti e dei mutualisti, si notano delle differenze. Ad esempio, all'interno dei parassiti, nessuna delle specie del genere *Rickettsia* (dal n. 3 al 7) ha come hub out il D-glucosio. Il confronto dei genomi di diversi endosimbionti (Klasson e Andersson, 2004) ha in effetti mostrato che *Rickettsia* spp. non possiedono gli enzimi del pathway glicolitico, per cui in queste specie il ruolo del glucosio non è centrale come nelle altre. Si pensa che acquisiscano glicerolo-3-fosfato dall'ospite: questo composto compare tra gli hub out (reagenti) di 3 specie di *Rickettsia* delle 5 analizzate in questo studio. Oppure, tutti i mutualisti, ad eccezione di *S. glossinidius*, non hanno succinato e acetilCoA come hub in (cioè tra i prodotti con maggiori legami). Le differenze riscontrate all'interno dei gruppi di parassiti e mutualisti potrebbero essere legate al carattere specie-specifico del processo di riduzione del genoma subito dagli organismi endosimbionti.

Conclusioni

L'analisi delle reti metaboliche di batteri endosimbionti e a vita libera ha permesso di individuare caratteristiche comuni a queste reti, ma anche differenze.

In particolare, tutte le reti metaboliche analizzate hanno una architettura di base comune: sono, cioè, delle reti a invarianza di scala. Questo è dimostrato dal fatto che la distribuzione dei gradi dei nodi segue una legge di potenza, cioè vi sono molti nodi con poche connessioni e pochi nodi, gli hub, con un numero elevato di link. Gli esponenti di grado della distribuzione sono molto simili per tutte le reti. Altre caratteristiche strutturali comuni alle reti studiate sono il diametro e la distanza media: entrambi i parametri si sono rivelati uniformi per tutte le reti. In particolare, la distanza media misurata è stata di $5,1-5,5$ per tutte le reti, ad indicare che possiedono la proprietà di “piccoli mondi”.

Seppure la topologia e gli aspetti strutturali sono comuni a tutte le reti, sono state individuate alcune differenze. In primo luogo le reti metaboliche hanno dimensioni diverse, misurate come numero di nodi. Gli endosimbionti tendono ad avere una rete più piccola rispetto ai batteri a vita libera. Confrontando questo dato con la dimensione del genoma si è visto che a genoma ridotto corrisponde una rete più piccola, anche se non si ha una riduzione lineare della rete in rapporto al genoma. Infatti un genoma piccolo produce, in proporzione, una rete 3 volte più grande rispetto ad un genoma di dimensioni maggiori.

Infine differenze si riscontrano negli hub: quelli principali sono gli stessi per tutte le reti, mentre quelli con grado minore tendono a variare in modo specie-specifico.

Bibliografia

ADAMIC, L.A., 1999. *Proceedings of the Third European Conference, ECDL'99* (Springer-Verlag, Berlin), p. 443.

AKMAN, L., AKSOY, S., 2001. A novel application of gene arrays: *Escherichia coli* array provides insight into the biology of the obligate endosymbiont of tsetse fly. *Proc. Natl. Acad. Sci. USA* **98**, 7546-7551.

AKMAN, L., YAMASHITA, A., WATANABE, H., OSHIMA, K., SHIBA, T., HATTORI, M., AKSOY, S., 2002. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nature genetics* **32**, 402-407.

AKSOY, S., RIO, R.V.M., 2005. Interaction among multiple genomes: Tsetse, its symbionts and trypanosomes. *Insect Biochemistry and Molecular Biology* **35**, 691-698.

ANDERSSON, G.E., ZOMORODIPOUR, A., ANDERSSON, J.O., SICHERITZ-PONTÉN, T., ET AL., 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133-143.

ALBERT, R., 2005. Scale-free networks in cell biology. *Journal of Cell Science* **118**, 4947-4957.

ALBERT, R. e BARABÁSI, A.-L., 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47-97.

ALBERT, R., JEONG, H. e BARABÁSI, A.-L., 1999. Diameter of the World Wide Web. *Nature* **401**, 130-131.

ALBERT, R., JEONG, H. e BARABÁSI, A.-L., 2000. Error and attack tolerance of complex networks. *Nature* **406**, 378-382.

ARITA, M., 2004. The metabolic world of *Escherichia coli* is not small. *Proc. Natl. Acad. Sci. USA* **101**, 1543-1547

BALDO, L., BARTOS, J.D., WERREN, J.H., BAZZOCCHI, C., CASIRAGHI, M., PANELLI, S., 2002. Different rates of nucleotide substitution in *Wolbachia* endosymbionts of arthropods and nematodes: arm races or host shifts? *Parassitologia* **44**, 179-187.

BANDI, C., ANDERSON, T.J.C., GENCHI, C., BLAXTER, M.L., 1998. Phylogeny of *Wolbachia* in filarial nematodes. *Proc R Soc Lond B* **265**, 2407-2413.

BARABÁSI, A.-L., 2004. *Link: la scienza delle reti*. Einaudi, Torino.

BARABÁSI, A.-L. e ALBERT, R., 1999. Emergence of scaling in random networks. *Science* **286**, 509-512.

BARABÁSI, A.-L., JEONG, H., RAVASZ, E., NÉDA, Z., SCHUBERT, A., VICSEK, T., 2001. Evolution of the social network of scientific collaborations. *Physica A* **311**, 580-614.

BARABÁSI, A.-L. e OLTVAI, Z. N., 2004. Network biology: understanding the cell's functional organization. *Nature reviews* **5**, 101-113.

- BATAGELJ, V., MRVAR, A., 1998. Pajek-program for large network analysis. *Connections* **21**, 47-57.
- BAZZOCCHI, C., CECILIANI, F., BRATTIG, N., 2001. I simbionti batterici (*Wolbachia*) delle filarie: implicazioni per il trattamento e la patologia delle filariosi. *Ann.Ist.Super.Sanità* **37**, 265-273.
- BEARD, C.B., O'NEILL, S.L., MASON, P., MANDELCO, L., WOESE, C.R., TESH, R.B., RICHARDS, F.F., AKSOY, S., 1993. Genetic transformation and phylogeny of bacterial symbionts from tsetse. *Insect Mol. Biol.* **1**, 123-131.
- BIANCONI, G., e BARABASI, A.-L., 2001. Competition and multiscaling in evolving networks, *Europhysics Letters* **54**, 436-442.
- BRATTIG, N.W., BAZZOCCHI, C., KIRSCHNING, C.J., REILING, N., BUTTNER, D.W., CECILIANI, F., GEISINGER, F., HOCHREIN, H., ERNST, M., WAGNER, H., BANDI, C., HOERAUF, A., 2004. The major surface protein of *Wolbachia* endosymbionts in filarial nematodes elicits immune responses through TLR2 and TLR4J. *Immunol* **173**, 437-445.
- BRAYTON, K.A., KAPPEMEYER, L.S., HERNDON, D.R., DARK, M.J., TIBBALS, D.L., PALMER, G.H., McGUIRE, T.C., KNOWLES, D.P., 2005. Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. *PNAS* **102**, 844-849.
- BUCHANAN, M., 2003. *Nexus*. Mondadori, Milano.
- BUCHANAN, M., 2001. *Ubiquità*. Mondadori, Milano.
- CASIRAGHI, M., BAIN, O., GUERRERO, R., MARTIN, C., PACACQUA, V., GARDNER, S.L., FRANCESCHI, A., BANDI, C., 2004. Mapping the presence of *Wolbachia pipientis* in the phylogeny of filarial nematodes: evidence for symbiont loss during evolution. *Int J Parasitol* **34**, 191-203.
- CHIU, C.-H., TANG, P., CHU, C., HU, S., BAO, Q., YU, J., CHOU, Y.-Y., WANG, H.-S., LEE, Y.-S., 2005. The genome sequence of *Salmonella enterica* serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. *Nucleic Acids Research* **33**, 1690-1698.
- DALE, C., WELBURN, S.C., 2001. The endosymbionts of tsetse flies: manipulating host-parasite interactions. *International Journal for Parasitology* **31**, 628-631.
- DE CASTRO, R. e GROSSMAN, J.W., 1999. Famous trails to Paul Erdős. *Mathematical Intelligencer* **21**, 51-63.
- DOUGLAS, A.E., 1988. Sulphate utilization in aphid symbiosis. *Insect. Biochem. Mol. Biol.* **18**, 599-605.
- DUNNING HOTOPP, J.C., LIN, M., MADUPU, R., CRABTREE, J., ANGIUOLI, S., EISEN, J., SESHADRI, R., REN, Q., WU, M., et al., 2006. Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genetics* **2**, 208-223.
- FERRER i CANCHO, E., SOLÈ, R.V., 2001. Santa Fe Institute working paper 01-03-016.

- FOSTER, J., GANATRA, M., KAMAL, I., WARE, J., MAKAROVA, K., IVANOVA, N., BHATTACHARYYA, A., KAPATRAL, V., KUMAR, S., et al., 2005. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *Plos Biol* **3**, e121.
- GIOT, L., BADER, J.S., BROUWER, C., CHAUDHURI, A., KUANG, B., LI, Y., HAO, Y.L., OOI, C.E., GODWIN, B., VITOLS, E. et al., 2003. A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727-1736.
- GIRIN, C., BOULETREAU, M., 1995. Microorganism-associated variation in host infestation efficiency in a parasitoid wasp *Trichogramma bourarachae*. *Experientia* **52**, 398-402.
- GOODNER, B., HINKLE, G., GATTUNG, S., MILLER, N., BLANCHARD, M., QUROLLO, B., GOLDMAN, B.S., et al., 2001. Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* **294**, 2323-2328.
- GRANOVETTER, M., 1983, The strenght of weak ties: a network theory revisited. *Sociological Theory* **1**, 201-233.
- HARTWELL, L.H., HOPFIELD, J.J., LEIBLER, S., MURRAY, A.W., 1999. From molecular to modular cell biology. *Nature* **402**, C47-C52.
- HOFFMAN, P., 1999. *Luomo che amava solo i numeri: la storia di Paul Erdős, un genio alla ricerca della verità matematica*. Mondadori, Milano.
- HURST, G.D.D., JIGGINS, F.M., VAN DER SCHULENBURG, J.H.D., BERTRAND, D., WEST, S.A., et al., 1999. Male-killing *Wolbachia* in two species of insect. *Proc. R. Soc. Londonn Ser. B.* **266**, 735-740.
- JEONG, H., MASON, S., BARABÁSI, A.-L., OLTVAI, Z. N., 2001. Centrality end lethality of protein networks. *Nature* **411**, 41-42.
- JEONG, H., OLTVAI, Z. N., BARABÁSI, A.-L., 2003. Prediction of protein essentiality based on genomic data. *ComplexUs* **1**, 19-28.
- JEONG, H., TOMBOR, B., ALBERT, R., OLTVAI, Z. N., BARABÁSI, A.-L., , 2000. The large-scale organization of metabolic networks. *Nature* **407**, 651-654.
- JIGGINS, F.M., HURST, G.D.D., YANG, Z., 2002. Host-symbiont conflicts: positive selection on an outer membrane protein of parasite but not mutualistic rickettsiaceae. *Mol Biol Evol* **19**, 1341-1349.
- KLASSON, L., ANDERSSON S.G.E., 2004. Evolution of minimal-gene-sets in host-dependent bacteria. *Trends in Microbiology* **12;1**, 37-42.
- LEVIN, S., 1999. *Fragile dominion: complexity and the commons*. Perseus, Cambridge (Mass.).
- LI, S., ARMSTRONG, C.M., BERTIN, N., GE, H., MILSTEIN, S., BOXEM, M., VIDALAIN, P.O., HAN, J.D., CHESNEAU, A., HAO, T. et al., 2004. A map of the interactome network of the metazoan *C.elegans*. *Science* **303**, 540-543.
- LO, N., CASIRAGHI, M., SALATI, E., BAZZOCCHI, C., BANDI, C., 2002. How many *Wolbachia* supergroups exist? *Mol Biol Evol* **19**, 341-346.

- MA, H., e ZENG, A.-P., 2003. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* **19**, 270-277.
- MALTSEV, N., GLASS, E., SULAKHE, D., RODRIGUEZ, A., SYED, M.H., BOMPADA, T., ZHANG, T., D'SOUZA, M., 2006. PUMA²-grid-based high-throughput analysis of genomes and metabolic pathways. *Nucleic Acid Res.* **1**; **34** (Database issue): D369-72.
- MATSUNAGA, T., OKAMURA, Y., FUKUDA, Y., WAHYUDI, A.T., MURASE, Y., TAKEYAMA, H., 2005. Complete genome sequence of the facultative anaerobic magnetotactic bacterium *Magnetospirillum* sp. strain AMB-1. *DNA research* **12**, 157-166.
- MAVROMATIS, K., DOYLE, C.K., LYKIDIS, A., IVANOVA, N., FRANCINO, M.P., CHAIN, P., SHIN, M., MALFATTI, S., LARIMER, F., COPELAND, A., et al., 2006. The genome of the obligately intracellular bacterium *Ehrlichia canis* reveals themes of complex membrane structure and immune evasion strategies. *Journal of Bacteriology* **188**, 4015-4023.
- McLEOD, M.P., QIN, X., KARPATY, E., GIOIA, J., HIGHLANDER, S.K., FOX, G.E., McNEILL, T.Z., JIANG, H., et al, 2004. Complete genome sequence of *Rickettsia typhi* and comparison with sequences of other Rickettsiae. *Journal of Bacteriology* **186**, 5842-5855.
- MIN, K.T., BENZER, S., 1997. Wolbachia, normally a symbiont of drosophila, can be virulent, causing degeneration and early death. *Proc. Natl. Acad. Sci. USA* **94**, 10792-96.
- MONTOYA, J.M., SOLÉ, R.V., 2000. Small world patterns in food web. *Cond-mat/0011195*.
- NEWMAN M.E.J., 1998, Models of the Small World. A review. *cond-mat/0001118*.
- NEWMAN, M., BARABÁSI, A.-L., WATTS, D., 2006. *The structure and dynamics of networks*. Princeton University Press, Princeton.
- NEWMAN, M.E.J., STROGATZ, S.H., WATTS, D.J., 2001. Random graphs with arbitrary degree distributions and their applications. *Physical review E* **64**, 026118
- OGATA, H., AUDIC, S., RENESTO-AUDIFFREN, P., FOURNIER, P-E., BARBE, V., SAMSON, D., et al., 2001. Mechanisms of evolution in *Rickettsia conorii* and *R.prowazekii*. *Science* **293**, 2093-2097.
- OGATA, H., LA SCOLA, B., AUDIC, S., RENESTO, P., BLANC, G., ROBERT, C., FOURNIER, P-E., CLAVERIE, J-M., RAOULT, D., 2006. Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens. *PLoS Genetics* **2**, 733-744
- OGATA, H., RENESTO, P., AUDIC, S., ROBERT, C., BLANC, G., FOURNIER, P-E., PARINELLO, H., CLAVERIE, J-M., RAOULT, D., 2005. The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biology* **3**, 1391-1402.

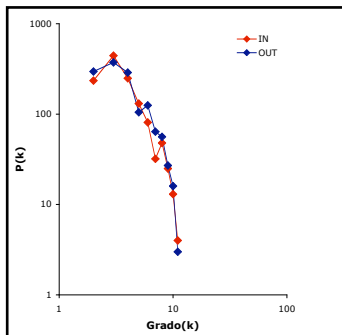
- PASTOR- SATORRAS, R., VASQUEZ, A., VESPIGNANI, A., 2001. Dynamical and correlation properties of the internet. *Phys. Rev. Lett.* **87**, 258701.
- PRŽULJ, N., 2004. Graph theory approaches to protein interaction data analysis. *Technical report 322/04*, Department of Computer Science, University of Toronto.
- RAIN, J.C., SELLG, L., DE REUSE, H., BATTAGLIA, V., REVERDY, C., SIMON, S., LENZEN, G., PETEL, F., WOJCIK, J., SCHACHTER, V., et al, 2001. The protein-protein interaction map of *Helicobacter pylori*. *Nature* **409**, 211-215.
- RAJAN, T.V., 2004. Relationship of anti-microbial activity of tetracyclines to their ability to block the L3 to L4 molt of the uman filarial parasite *Brugia malayi*. *Am J Trop Med Hyg* **71**, 24-28.
- ROUSSET, F., BOUCHON, D., PINTUREAU, B., JUCHAULT, P., SOLIGNAC, M., 1992. *Wolbachia* endosymbionts responsible for various alterantions of sexuality in arthropods. *Proc. R. Soc. London Ser. B* **250**, 91-98.
- SHIGENOBU, S., WATANABE, H., HATTORI, M., SAKAKI, Y., ISHIKAWA, H., 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp.APS. *Nature* **407**, 81-86.
- SOLÉ, R.V., PASTOR-SATORRAS, R., SMITH, E. E KEPLER, T.B., 2002. A model of large-scale proteome evolution. *Advances in Complex Systems* **5**, 43-54.
- STOUTHAMER, R., BREEUWER, J.A.J., HURST, G.D.D., 1999. *Wolbachia* pipientis: Microbial manipulator of Arthropod reproduction. *Ann. Rev. Microbiol.* **53**, 71-102.
- STOUTHAMER, R., LUCK, R.F., HAMILTON. W.D., 1990. Antibiotics cause parthenogenetic *Trichogramma* to revert to sex. *Proc. Natl. Acad. Sci. USA* **87**, 2424-27.
- TAMAS, I., KLASSON, L., CANBÄCK, B., NÄSLUND, A.K., ERIKSSON, A-S., WERNEGREEN, J.J., SANDSTRÖM, J.P., MOEAN, N.A., ANDERSSON, G.E., 2002. 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296**, 2376-2379.
- TOH, H., WEISS, B.L., PERKIN, S.A.H., YAMASHITA, A., OSHIMA, K., HATTORI, M., AKSOY, S., 2006. Massive genome erosion and functional adaptation provide insight into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome research* **16**, 149-156.
- TRAVERS, J. e MILGRAM, S., 1969. An experimental study of the small world problem. *Sociometry* **32**, 425-443.
- VANDEKERCKOVE, T.T.M., WATTEYNE, S., WILLEMS, A., SWING, J.G., MERTENS, J., GILLIS, M., 1999. Phylogenetic analysis of the 16S rDNA of the cytoplasmic bacterium *Wolbachia* from the novel host *Folsomia candida* (Hexapoda, Collembola) and its implications for wolbachial taxonomy. *FEMS Microbiol Lett* **180**, 279-286.
- VAN HAM, R.C.H.J., KAMERBEEK, J., PALACIOS, C., RAUSELL, C., ABA-SCAL., F., BASTOLLA, U., FERNANDEZ, J.M., JIMENEZ, L., POSTIGO, M., SILVA, F.J., et al., 2002. Reductive genome evolution in *Buchnera aphidicola*. *PNAS* **100**, no.2, 581-586.

- VÁSQUEZ, A., FLAMMINI, A., MARITAN, A., e VESPIGNANI, A., 2003. Modeling of protein interaction networks. *Complexus* **1**, 38-44.
- VOGELSTAIN, B., LANE, D., LEVINE, A.J., 2000, Surfing the p53 network. *Nature* **408**, 307-310.
- WAGNER, A. e FELL, D., 2001. The small world inside large metabolic networks. *Proc. R. Soc. London B* **268**, 1803-1810.
- WATTS, D. e STROGATZ, S.H., 1998. Collective dynamics of 'small-world' networks. *Nature* **393**, 440-442.
- WERNEGREEN, J.J., 2002. Genome evolution in bacterial endosymbionts of insect. *Nat. Rev. Genet.* **3**, 850-861.
- WERREN, J., 1997. Biology of *Wolbachia*. *Annu. Rev. Entomol.* **42**, 587-609.
- WERREN, J.H., ZHANG, W., GUO, L.R., 1995. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proc R Soc Lond B* **261**, 55-63.
- WOOD, D.W., SETUBAL, J.C., KAUL, R., MONKS, D.E., KITAJIMA, J.P., OKURA, V.K., ZHOU, Y., CHEN, L., WOOD, G.E., ALMEIDA Jr, N.F., WOO, L., CHEN, Y., et al., 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* **294**, 2317-2323.
- WU, M., SUN, L.V., VARNATHEVAN, J., RIEGLER, M., DEBOY, R., BROWNLIE, J.C., MCGRAW, E.A., MARTIN, W., ESSER, C., AHMADI-NEJAD, N., et al., 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PlosBiol* **2**, 327-341.
- YOOK, S., JEONG, H., BARABÁSI, A.-L., 2001a. Weighted evolving networks. *Phys. Rev. Lett.* **86**, 5835-5838.

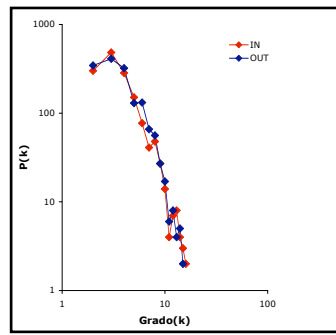
APPENDICI

APPENDICE A

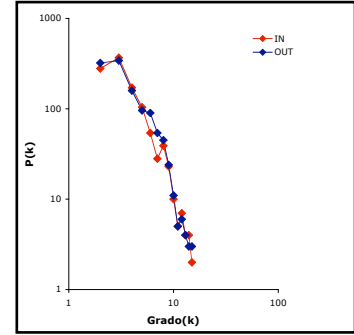
Distribuzione dei gradi dei nodi, in e out, per tutte le reti analizzate.



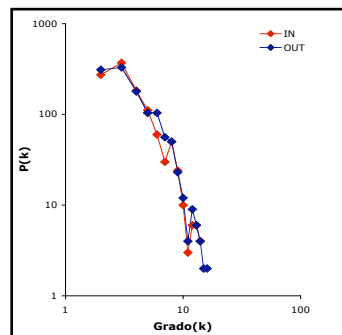
Anaplasma phagocytophilum



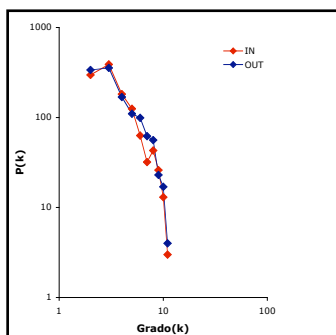
Anaplasma marginale



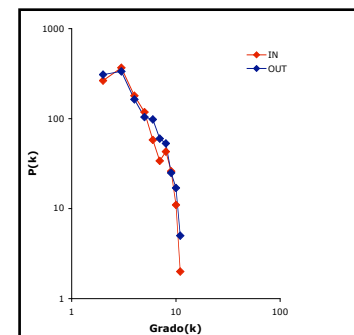
Rickettsia typhi



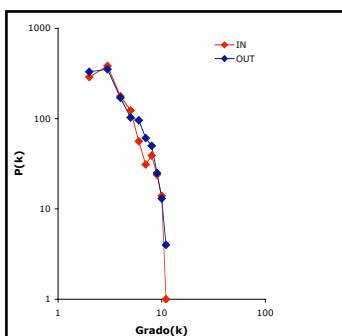
Rickettsia prowazekii



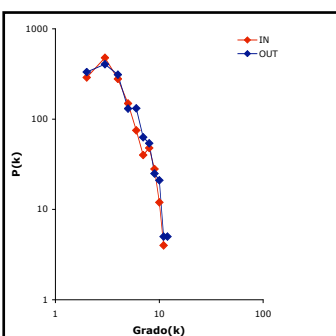
Rickettsia felis



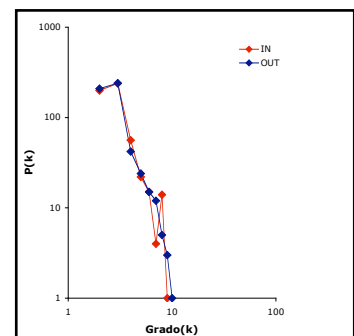
Rickettsia conorii



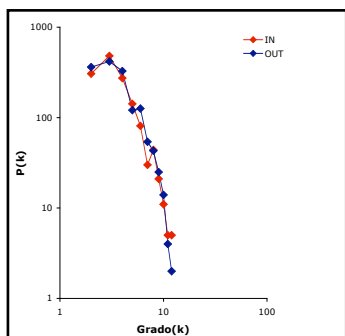
Rickettsia bellii



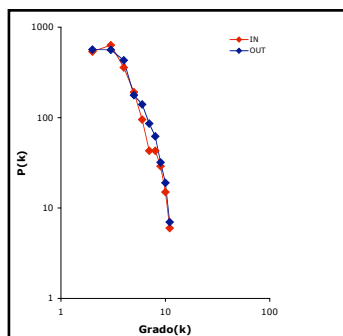
Ehrlichia canis



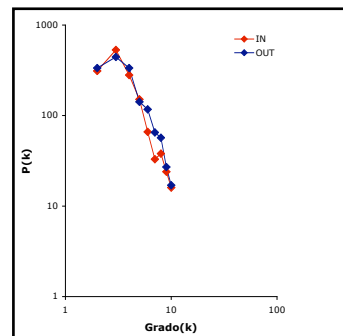
Wolbachia pipientis



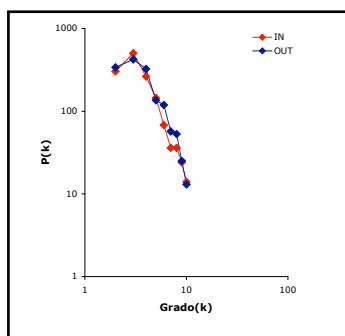
Wigglesworthia glossinidia



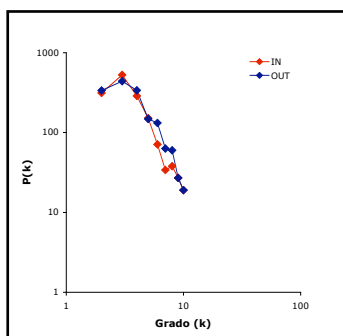
Sodalis glossinidius



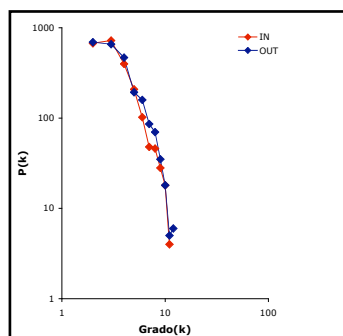
Buchnera aphidicola str. Sg



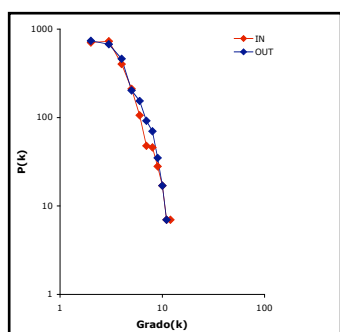
Buchnera aphidicola str. Bp



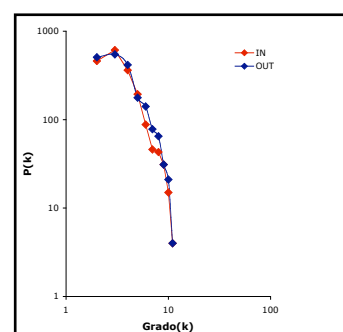
Buchnera aphidicola str. Ap



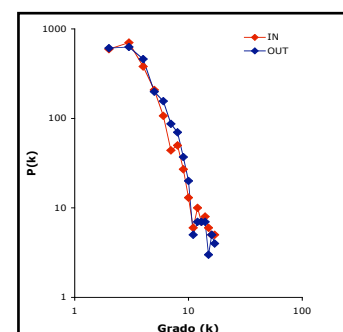
Salmonella enterica



Escherichia coli



Magnetospirillum magneticum



Agrobacterium tumefaciens

APPENDICE B

Lista degli hub individuati per tutte le reti metaboliche analizzate.

Anaplasma phagocytophilum

Hub in	Grado
ADP	357
H ₂ O	329
ortofosfato	325
ATP	213
CO ₂	169
NADH	125
Piruvato	119
pirofosfato	103
FADH ₂	55
S-lattato	54
2-oxoglutarato	43
CoA	42
NADP ⁺	41
Glutazione	30
L-glutamato	29
acetilCoA	28
Succinato	28

Hub out	Grado
ATP	381
H ₂ O	313
ADP	208
ortofosfato	197
D-glucosio	138
NAD ⁺	134
D-glucosio	65
FAD	41
L-glutamato	40
Lipoammide	37
Pcarrier-prot-Phis	32
acetilCoA	31
NADPH	30
Glutazione	28
CoA	26
Piruvato	25
Glicina	23

Anaplasma marginale

Hub in	Grado
ADP	396
ortofosfato	347
H ₂ O	275
ATP	213
CO ₂	175
NADH	133
Piruvato	129
pirofosfato	120
AMP	84
FADH ₂	64
2-oxoglutarato	58
S-lattato	56
NADP ⁺	55
CoA	47
Nh ₃	41
L-glutamato	38
Diidrolipoammide	37
NAD ⁺	35
Glutazione	31

Hub out	Grado
ATP	479
H ₂ O	356
ADP	207
ortofosfato	199
NAD ⁺	142
D-glucosio	138
L-glutamato	53
FAD	50
Piruvato	43
Alfa-D-glucosio	39
acetilCoA	38
L-aspartato	37
Lipoammide	37
NADPH	36
Pcarrier-prot-Phis	32
CoA	29
Glutazione	28
2-oxoglutarato	25
L-glutamina	25
Glicerolo-3P	24

Rickettsia typhi

Hub in	Grado
ADP	306
ortofosfato	272
CO ₂	153
pirofosfato	88
H ₂ O	79
AMP	61
2-oxoglutarato	56
NADH	55
CoA	47
FADH ₂	46
ATP	42
NADP ⁺	40
Piruvato	38
Diidrolipoammide	37
NH ₃	37
acetilCoA	30
Succinato	26
CMP	26
Acetato	25
NADPH	25

Hub out	Grado
ATP	325
H ₂ O	324
L-glutamato	53
FAD	42
ADP	40
Piruvato	40
Lipoammide	37
NAD ⁺	35
acetilCoA	35
NADPH	35
CoA	31
Glicerolo-3P	24
L-aspartato	24
Ortofosfato	24
Oxaloacetato	22

Rickettsia prowazekii

Hub in	Grado
ADP	301
ortofosfato	269
CO ₂	170
H ₂ O	108
pirofosfato	94
NADH	86
AMP	61
2-oxoglutarato	56
ATP	54
CoA	48
Piruvato	47
FADH ₂	44
NADP ⁺	41
Diidrolipoammide	37
NH ₃	34
acetilCoA	29
Succinato	27

Hub out	Grado
H ₂ O	328
ATP	327
ADP	54
L-glutamato	53
FAD	42
Piruvato	40
Lipoammide	37
NAD ⁺	37
acetilCoA	36
NADPH	35
CoA	32
L-aspartato	25
NADP ⁺	24
S-malato	24
Co ₂	24
Succinil-CoA	22

Rickettsia felis

Hub in	Grado
ADP	312
ortofosfato	304
CO ₂	179
pirofosfato	93
H ₂ O	81
AMP	65
2-oxoglutarato	62
Nadh	58
CoA	54
NADP+	52
ATP	49
Succinato	46
FADH ₂	46
Piruvato	44
Nh ₃	44
Diidrolipoammide	37
L-glutamato	32
Acetato	30
AcetilCoA	30

Hub out	Grado
ATP	373
H ₂ O	361
L-glutamato	53
ADP	45
acetilCoA	42
FAD	42
Piruvato	40
NAD+	38
Lipoammide	37
NADPH	36
2-oxoglutarato	35
CoA	31
L-aspartato	27
Glicerolo-3P	24
Ortofosfato	24

Rickettsia conorii

Hub in	Grado
ADP	307
ortofosfato	268
CO ₂	174
pirofosfato	90
NADH	81
H ₂ O	79
AMP	63
2-oxoglutarato	62
CoA	54
NADP+	50
Succinato	46
FADH ₂	46
ATP	44
Nh ₃	43
Piruvato	42
Diidrolipoammide	37
L-glutamato	32
NADPH	26

Hub out	Grado
H ₂ O	341
ATP	330
L-glutamato	53
ADP	42
acetilCoA	42
FAD	42
Piruvato	40
NAD+	37
Lipoammide	37
NADPH	36
2-oxoglutarato	35
CoA	31
L-aspartato	27
Glicerolo3P	24
Oxaloacetato	24

Rickettsia bellii

Hub in	Grado
ADP	306
Ortofosfato	273
Co2	180
Pirofosfato	90
H2O	83
AMP	63
2-oxoglutarato	62
NADH	56
CoA	53
NADP+	52
Atp	49
Succinato	46
FADH2	46
Piruvato	44
Nh3	44
Diidrolipoammide	37
L-glutamato	32
Acetato	30
AcetilCoA	29
NADPH	26

Hub out	Grado
ATP	351
H2O	348
L-glutamato	53
Ortofosfato	46
ADP	45
acetilCoA	41
Piruvato	40
FAD	40
NAD+	38
Lipoammide	37
NADPH	36
2-oxoglutarato	35
CoA	29
L-aspartato	27
Glicerolo-3P	24

Ehrlichia canis

Hub in	Grado
ADP	420
ortofosfato	305
H2O	272
ATP	216
CO2	169
NADH	137
pirofosfato	117
Piruvato	116
AMP	84
FADH2	65
S-lattato	54
2-oxoglutarato	49
NADP+	48
CoA	43
Diidrolipoammide	37
Nh3	34
Nad+	32
AcetilCoA	30

Hub out	Grado
ATP	446
H2O	354
ADP	210
ortofosfato	202
NAD+	146
D-glucosio	138
FAD	51
L-glutamato	47
L-aspartato	40
Lipoammide	37
Alfa-D-glucosio	37
acetilCoA	33
Pcarrier-protein	32
Piruvato	31
CoA	29
NADPH	28
Glutatione	28
Glicina	25
Glicerolo3P	24
2-oxoglutarato	24

Wigglesworthia glossinidia

Hub in	Grado
ADP	389
ortofosfato	338
H ₂ O	269
ATP	218
CO ₂	151
pirofosfato	133
Piruvato	125
NADH	116
AMP	71
S-lattato	56
FADH ₂	51
NAD ⁺	49
NADP ⁺	40
2-oxoglutarato	38
L-glutamato	37
Diidrolipoammide	34
Glutazione	30
NH ₃	28

Hub out	Grado
ATP	426
H ₂ O	350
ADP	212
ortofosfato	201
D-glucosio	138
NAD ⁺	98
L-glutamato	45
Alfa-D-glucosio	39
FAD	39
L-aspartato	35
Lipoammide	34
Pcarrier-proteine	32
Glutazione	28
L-glutamina	27
NADPH	26
Piruvato	25
Glicerolo ₃ P	24

Sodalis glossinidius

Hub in	Grado
ADP	482
ortofosfato	388
H ₂ O	339
Co ₂	260
ATP	219
pirofosfato	199
Piruvato	184
NADH	106
Tioredoxina	95
2-oxoglutarato	94
NAD ⁺	87
Nh ₃	80
NADP ⁺	71
FADH ₂	71
L-glutamato	64
CoA	58
S-lattato	56
D-gliceraldeide-3P	50
Acetato	50
Succinato	48
NADPH	37
Diidrolipoammide	36
AcetilCoA	33

Hub out	Grado
ATP	560
H ₂ O	513
ortofosfato	221
ADP	212
NAD ⁺	187
D-glucosio	145
L-glutamato	78
Acetil-CoA	55
NADPH	49
FAD	47
L-aspartato	45
2-oxoglutarato	44
Chorismate	43
Piruvato	41
Alfa-D-glucosio	39
L-glutamina	36
Lipoammide	36
5P-alfa-D-ribosio-1diP	32
O ₂	31
L-arginina	29
Glutazione	28
CoA	27
Co ₂	26

Buchnera aphidicola str. Sg

Hub in	Grado
ADP	382
ortofosfato	330
H ₂ O	289
ATP	210
CO ₂	185
pirofosfato	162
Piruvato	139
NADH	124
AMP	86
2-oxoglutarato	61
S-lattato	56
FADH ₂	52
NH ₃	42
acetato	42
CoA	38
L-glutamato	34
diidrolipoammide	34
D-glicer aldeide-3P	31
Glutazione	30
NAD ⁺	30

Hub out	Grado
ATP	427
H ₂ O	368
ortofosfato	209
ADP	203
D-glucosio	140
NAD ⁺	129
NADPH	61
L-glutamato	54
Alfa-D-glucosio	39
L-aspartato	38
chorismate	37
lipoammide	34
FAD	34
Acetil-CoA	34
piruvato	33
Pcarrier-protein	32
5P-alfa-D-ribosio-1-diP	31
glutazione	28
L-serina	23
2-oxoglutarato	22

Buchnera aphidicola str. Bp

Hub in	Grado
ADP	370
H ₂ O	283
ortofosfato	265
ATP	210
CO ₂	172
Piruvato	142
pirofosfato	118
AMP	74
NADH	66
2-oxoglutarato	59
S-lattato	56
NADP ⁺	51
FADH ₂	48
NH ₃	44
acetato	41
diidrolipoammide	34
CoA	33
L-glutamato	32
D-glicer aldeide3P	31
Glutazione	30

Hub out	Grado
ATP	393
H ₂ O	361
ortofosfato	206
ADP	203
D-glucosio	140
NAD ⁺	131
L-glutamato	57
chorismate	43
Alfa-D-glucosio	39
L-aspartato	37
NADPH	35
lipoammide	34
FAD	34
Acetil-CoA	33
piruvato	33
Pcarrier-protein	32
L-glutammina	31
glutazione	28

Buchnera aphidicola str. Ap

Salmonella enterica

Hub in	Grado
ADP	383
ortofosfato	331
H ₂ O	289
ATP	213
CO ₂	192
pirofosfato	163
piruvato	143
NADH	130
AMP	86
2-oxoglutarato	64
NADP+	62
2-lattato	56
FADH ₂	51
CoA	41
NH ₃	41
L-glutamato	38
ferricromo	38
diidrolipoammide	37
D-gliceraldeide-3P	31
NAD+	30
Glutazione	30

Hub out	Grado
ATP	428
H ₂ O	372
ortofosfato	212
ADP	206
D-glucosio	140
NAD+	136
NADPH	61
L-glutamato	55
chorismate	43
Alfa-D-glucosio	39
L-aspartato	38
lipoammide	37
FAD	37
Acetil-CoA	37
piruvato	34
Pcarrier-protein	32
L-glutamina	32
5P-alfa-D-ribosio-1diP	31
glutazione	28
L-serina	23

Hub in	Grado
ADP	531
ortofosfato	433
H ₂ O	368
Co ₂	304
NADH	247
ATP	224
Piruvato	208
Pirofosfato	207
Nh ₃	125
NAD+	108
AMP	101
NADPH	100
2-oxoglutarato	94
NADP+	81
L-glutamato	75
FADH ₂	75
D-gliceraldeide-3P	63
Acetato	57
S-lattato	56
Succinato	56
AcetilCoA	42
H ₂ O ₂	37
Diidrolipoammide	37
H+	35

Hub out	Grado
H ₂ O	655
ATP	476
NAD+	231
Ortofosfato	227
ADP	214
D-glucosio	188
NADP+	148
CoA	85
L-glutamato	78
acetilCoA	63
NADH	57
FAD	54
D-pinitolo	52
proteina_III (Mtl)_histidine	51
O ₂	48
Piruvato	47
Butanolo	46
Chorismate	43
L-glutamina	40
Lipoammide	37
5P-α-D-ribosio-1diP	33
Pcarrier-protein	32
NAD+	30
L-arginina	29
Co ₂	28
Glutazione	28
NADPH	27

Magnetospirillum magneticum

Escherichia coli

Hub in	Grado
ADP	587
ortofosfato	396
H ₂ O	330
CO ₂	271
ATP	224
Piruvato	171
NADH	169
AMP	91
2-oxoglutarato	80
NAD ⁺	76
CoA	69
NADP ⁺	68
Nh ₃	68
L-glutamato	64
FADH ₂	63
Acetato	57
S-lattato	56
Succinato	49
AcetilCoA	43
Diidrolipoammide	37
NADPH	35
Pirofosfato	33
Glutazione	31

Hub out	Grado
ATP	511
H ₂ O	497
ADP	214
ortofosfato	212
D-glucosio	141
NAD ⁺	136
L-glutamato	65
AcetilCoA	58
Piruvato	54
FAD	52
NADPH	45
2-oxoglutarato	45
L-aspartato	41
Alfa-D-glucosio	39
CoA	38
Chorismate	37
L-glutammina	36
O ₂	36
Pcarrier-protein	32
Glutazione	28
Co ₂	27
L-arginina	27
P-α-D-ribosio-1-diP	25
Glicina	25
glicerolo ₃ P	24
NADH	24

Hub in	Grado
ADP	534
ortofosfato	438
H ₂ O	370
CO ₂	314
NADH	257
ATP	227
Piruvato	215
pirofosfato	210
NH ₃	121
AMP	101
CoA	100
2-oxoglutarato	94
NADP ⁺	88
FADH ₂	81
NADPH	73
L-glutamato	72
D-gliceraldeide-3P	63
acetato	58
S-lattato	56
AcetilCoA	42
AcetilCoA	42
H ⁺	38
Diidrolipoammide	37

Hub out	Grado
ATP	627
H ₂ O	637
NAD ⁺	239
ortofosfato	232
ADP	217
D-glucosio	150
L-glutamato	76
acetilCoA	62
FAD	54
2-oxoglutarato	51
NADPH	85
Piruvato	48
L-aspartato	46
Chorismate	43
O ₂	41
L-glutammina	40
Lipoammide	37
CoA	59
5P-α-D-ribosio-1-diP	33
Pcarrier-protein	32
L-arginina	29
Co ₂	28
NADH	28
Glutazione	28

Agrobacterium tumefaciens

Hub in	Grado
ADP	509
ortofosfato	412
H ₂ O	403
CO ₂	303
ATP	220
NADH	216
Piruvato	210
pirofosfato	198
Nh ₃	110
NAD ⁺	102
AMP	94
2-oxoglutarato	94
NADP ⁺	78
FADH ₂	75
L-glutamato	74
NADPH	72
CoA	67
D-gliceraldeide-3P	63
S-lattato	56
Succinato	51
Acetato	49
H ₂ O ₂	42
AcetilCoA	41
Diidrolipoammide	37

Hub out	Grado
ATP	596
H ₂ O	568
ortofosfato	226
ADP	216
NAD ⁺	179
D-glucosio	149
L-glutamato	74
CoA	56
FAD	54
2-oxoglutarato	53
O ₂	52
NADPH	51
Piruvato	48
NADP	46
Nh ₃	44
Chorismate	43
Alfa-D-glucosio	39
L-glutammina	38
Lipoammide	37
5P-alfa-D-ribosio-1diP	32
Pcarrier-protein	32
NADH	30
L-arginina	28
Glutatione	28
Glicina	25
Co ₂	25

