

UNIVERSITA' DEGLI STUDI DI MILANO - BICOCCA

Facoltà di Scienze Matematiche, Fisiche e Naturali

Corso di Laurea Specialistica in Biologia



## LA TEORIA DELLE RETI NEGLI STUDI GENOMICI

Relatore: Dott. Maurizio CASIRAGHI

Tesi di Laurea di:

Cinzia MORETTI

Matr. 048040

Anno Accademico 2006 – 2007

# INDICE

SINTESI .....	3
INTRODUZIONE .....	5
1) LA SCIENZA DELLA COMPLESSITÀ E LA TEORIA DELLE RETI .....	6
1. Introduzione alla teoria delle reti .....	8
2. LE RETI DELLA VITA .....	9
2.1. Le reti sociali .....	9
2.2. Le reti ecologiche .....	12
2.3. Le reti cellulari .....	12
3. CENNI STORICI: dalla teoria dei grafi alla nuova scienza delle reti .....	13
3.1. La teoria dei grafi .....	13
3.1.1. L'effetto piccolo-mondo .....	16
3.2. Le reti piccolo-mondo tra ordine e casualità .....	17
3.3. Gli <i>hub</i> e le reti a invarianza di scala .....	18
4. I MODELLI MATEMATICI PER LO STUDIO DELLE RETI .....	20
4.1. I grafi casuali di Erdős e Rényi .....	20
4.2. Le reti piccolo-mondo di Watts e Strogatz .....	22
4.3. La legge di potenza e le reti a invarianza di scala di Barabási e Albert ...	25
4.4. Modello a fitness di Bianconi e Barabási .....	28
5. LE RETI METABOLICHE .....	29
2) GLI ORGANISMI ANALIZZATI .....	33
1. L'endosimbiosi .....	33
2. Batteri endosimbionti intracellulari .....	36
2.1. Ordine Rickettsiales .....	36
2.2. <i>Wolbachia pipientis</i> .....	36
2.3. Genere <i>Rickettsia</i> .....	41
2.4. <i>Ehrlichia chaffeensis</i> .....	43

3. Batteri endosimbionti a vita libera .....	44
4. Gli ospiti .....	46
4.1. Le filarie .....	49
3) METODI .....	53
1. Costruzione delle reti metaboliche .....	53
1. 1. Il database .....	53
1. 2. Costruzione della rete metabolica .....	53
2. Analisi delle reti metaboliche .....	54
2. 1. Distribuzione dei gradi dei nodi .....	55
2. 2. Diametro e distanza media tra i nodi .....	55
2. 3. Hub .....	55
3. Identificazione delle molecole di interazione .....	55
3. 1. Analisi dell'interazione ospite – simbiote .....	56
3. 2. Indici di somiglianza .....	56
3. 3. Costruzione dell'albero filogenetico a partire dalle liste di 'solo prodotti' e 'solo reagenti' .....	57
4. Costruzione dell'albero filogenetico .....	58
4) RISULTATI E DISCUSSIONE .....	59
1. Analisi delle reti metaboliche .....	59
2. Analisi dell'interazione ospite-simbiote .....	65
2. 1. Interazione <i>Wolbachia pipientis</i> – nematode .....	66
2. 2. Molecole di interazione negli organismi patogeni .....	73
2. 3. Indici di somiglianza e alberi .....	76
5) CONCLUSIONI .....	80
6) BIBLIOGRAFIA .....	82
7) APPENDICI .....	87

## SINTESI

‘Le reti sono ovunque’: in qualsiasi momento, noi siamo il risultato ma anche parte integrante di svariate reti a partire dal livello microscopico fino a reti che avvolgono l’intero globo terrestre. Matematicamente la rete è un grafo in cui degli oggetti (nodi) sono collegati tra loro da legami (*link*) di varia natura.

E’ stato dimostrato che attraverso la nostra rete di conoscenze è possibile raggiungere qualsiasi altra persona del mondo con, in media, solo 6 passaggi (o ‘6 strette di mano’ nella visione comunemente nota come “i sei gradi di separazione”); nel web è possibile passare da una qualunque pagina web ad un’altra scelta a caso con in media 11 *click*; la media dei gradi di separazione tra le specie all’interno della rete alimentare, tra i neuroni nella rete neurale del nematode *Caenorhabditis elegans*, tra le molecole all’interno della cellula varia tra 2 e 14. Queste reti, pur apparentemente così diverse tra loro, posseggono caratteristiche comuni e sono definite ‘piccolo-mondo’ perché possiedono una elevata connettività: è possibile passare da un nodo all’altro con un numero molto basso di passaggi. Le reti piccolo-mondo e molte altre fanno parte della categoria dei sistemi complessi in quanto sono composti da molti elementi interagenti il cui comportamento globale non è comprensibile né prevedibile dallo studio delle singole parti.

La disciplina che si occupa di comprendere questi fenomeni è la *teoria delle reti* che rappresenta uno dei più recenti sviluppi della scienza della complessità e coinvolge discipline molto distanti tra loro come fisica, sociologia, biologia, informatica, economia. La teoria delle reti si pone come obiettivo quello di sviluppare nuovi approcci e tecniche per comprendere i sistemi naturali e artificiali; per esempio si è scoperto che attraverso le reti è possibile comprendere la dinamica di molti fenomeni per molto tempo ritenuti troppo complessi o complicati per poter essere spiegati: la diffusione delle epidemie, la crescita di Internet o i processi di comunicazione, per citarne solo alcuni.

Lo studio delle reti pone le sue radici nella ben più antica teoria dei grafi fondata tre secoli fa da Eulero, che per primo rappresentò una situazione reale tramite un grafo, cioè un insieme di nodi connessi da link. Da allora vari modelli sono stati proposti per spiegare e rappresentare le reti reali: per primi furono due matematici ungheresi, Erdős e Rényi che si accorsero che i grafi reali non sono affatto regolari, ma estremamente complicati ed equipararono la complessità alla casualità. Ma l’esperienza personale di ciascuno e la sociologia, una delle prime discipline che iniziò a ‘pensare a rete’, contraddicono la loro proposta: il modello casuale non coglie molte delle caratteristiche delle reti sociali come per esempio la vicinanza geografica tra le persone o il tipo di legame che si instaura tra i nodi.

Il primo esperimento applicato alle reti sociali risale alla fine degli anni Sessanta, fu realizzato dal sociologo Stanley Milgram e portò alla scoperta che è sufficiente una catena di sei persone (i già citati “sei gradi di separazione”) per mettere in relazione un individuo qualsiasi sul pianeta con qualunque altro. Il sociologo Mark Granovetter nel 1973 dimostrò che in una rete sociale i legami più importanti che mantengono la coerenza di una rete sono paradossalmente i legami deboli, cioè le conoscenze più superficiali che fungono da “ponti” tra gruppi di amici. Furono Watts e Strogatz che nel 1998 collegarono ingegnosamente i due lavori: scoprirono che i nodi si possono collegare in maniera né ordinata né casuale proponendo un nuovo modello di rete che si trova a metà tra i due estremi: partendo da una rete regolare e, aggiungendo delle connessioni casuali, i due ricercatori ottennero un nuovo modello di rete denominato “piccolo-mondo”. Le reti piccolo-mondo sono caratterizzate da un basso grado di separazione, cioè ogni coppia di nodi è collegata attraverso un numero minimo di passaggi, e da un alto coefficiente di *clustering* cioè ci sono piccoli gruppi di nodi altamente connessi al loro interno e allo stesso tempo collegati agli altri *cluster* da legami deboli.

Infine grazie all’accessibilità a Internet e a numerosi studi condotti con lo scopo di mappare il web è stato possibile scoprire un nuovo modello di rete più realistico: le “reti a invarianza di scala” (Barabási e Albert, 1999). In queste reti esistono nodi altamente connessi (*hub*) che acquisiscono link con maggiore rapidità di quelli che hanno pochi collegamenti, determinando una ripartizione delle connessioni distribuita secondo una legge di potenza. Questo tipo di dinamica sembra essere particolare ed esclusiva dei sistemi complessi, anche di quelli biologici. Nelle reti metaboliche, oggetto di studio di questo lavoro, i nodi rappresentano i substrati e i link le reazioni; contando il numero di substrati con ugual numero di connessioni si ottiene una legge di potenza, con pochi nodi che hanno molte connessioni e molti nodi con un basso numero di link.

La mia tesi si inserisce in una linea di ricerca volta allo studio delle simbiosi tra metazoi e microrganismi attraverso la teoria delle reti. In particolare, l’obiettivo della presente tesi è quello di studiare le reti metaboliche di diversi microrganismi che hanno instaurato con l’ospite diversi tipi di simbiosi. Particolare attenzione è stata rivolta alle simbiosi parassitarie, per indagare la relazione ospite-parassita, nei casi particolari di batteri che provocano patologie nell’uomo: modello principale delle analisi è stato il batterio *Wolbachia pipientis*, simbionte intracellulare di nematodi filaridi.

## INTRODUZIONE

La teoria delle reti permette di rappresentare attraverso nodi connessi da *link* gran parte dei sistemi reali. In questo lavoro le reti metaboliche di diversi microrganismi sono state studiate mediante questo approccio: i nodi rappresentano i substrati e i prodotti collegati tra loro da *link* che corrispondono alla reazione chimica che li coinvolge. Le reti metaboliche sono state ottenute a partire dai dati provenienti da progetti di sequenziamento genomico disponibili in database on-line.

In particolare l'obiettivo della presente tesi è stato quello di studiare il metabolismo di batteri che hanno instaurato con l'ospite diversi tipi di simbiosi ponendo l'accento sulle simbiosi parassitarie con batteri che provocano patologie nell'uomo.

È stata analizzata la topologia delle reti costruite per verificare a quale dei vari modelli proposti dalla teoria sono riconducibili; data la loro particolare struttura è stato inoltre possibile identificare le molecole chiave del metabolismo e quelle che sono oggetto di scambio nell'interazione ospite-simbionte.

Modello principale delle analisi è stato il batterio *Wolbachia pipientis* che ricopre contemporaneamente un doppio ruolo: di simbiote mutualista intracellulare nei nematodi filaridi e agente patogeno nell'uomo nelle filariosi. Infatti gli studi condotti fin'ora hanno evidenziato l'importante ruolo di *W. pipientis* nelle patologie causate dalle filarie e quindi la presenza del batterio è un aspetto della biologia delle filarie che può essere sfruttato per sviluppare un nuovo approccio per il controllo e la cura delle filariosi. Si è indagata in modo più approfondito la stretta relazione che lega *W. pipientis* e l'ospite nematode *Brugia malayi*; è stata perciò costruita la rete metabolica di *W. pipientis* endosimbionte del nematode parassita (*wBm*) e confrontata con quella di un altro nematode, *Caenorhabditis elegans*. La scelta di *C. elegans* è dovuta al fatto che la rete metabolica di *B. malayi* non è ancora disponibile nel database, dato che i lavori sul sequenziamento del suo genoma stanno dando ora i primi risultati (Ghedini *et al.*, 2007); è stata costruita perciò la rete metabolica di *C. elegans*, organismo modello molto ben conosciuto. Sono stati inoltre analizzate e confrontate le reti metaboliche di altri microrganismi: gli agenti patogeni del tifo (*Rickettsia prowazekii* e *R. typhi*), dell'erlichiosi monocitica umana (*Ehrlichia chaffeensis*), e due batteri patogeni extracellulari, *Escherichia coli* e *Salmonella enterica*.

Il metodo qui proposto ha permesso di studiare attraverso il nuovo approccio delle reti la simbiosi, individuando quelle molecole che probabilmente sono oggetto di scambio tra gli organismi. Dal confronto dei risultati ottenuti con le informazioni disponibili dalla bibliografia è stato possibile supportare questo approccio e individuare ulteriori potenzialità.

## 1) LA SCIENZA DELLA COMPLESSITÀ E LA TEORIA DELLE RETI

Lo studio delle reti rappresenta uno dei più recenti sviluppi della scienza della complessità, argomento che negli ultimi anni sta assumendo un'importanza trasversale per molte discipline apparentemente non connesse tra loro. I contributi giungono da ambiti molto disparati: fisica, matematica, informatica, biologia, filosofia, letteratura, economia, ecc. La scienza della complessità nasce alla fine degli anni Sessanta in seguito alla constatazione che l'approccio cartesiano non è più adeguato a spiegare il mondo. Il paradigma riduzionista, con il suo processo di 'smontaggio' e il motto *divide et impera*, infatti, si è rivelato utile fino a un certo punto.

La teoria della complessità, invece, si pone come superamento del riduzionismo e si orienta verso una visione più olistica, secondo cui il mondo è un insieme integrato, una rete di fenomeni fondamentalmente interconnessi e interdipendenti in cui le proprietà derivano dalle relazioni tra le sue parti; per usare una famosa formulazione: "il tutto è maggiore della somma delle sue parti". In realtà, appare sempre più chiaro che per la comprensione dei sistemi reali non si può ridurre, né il tutto alle parti, né le parti al tutto, ma è necessario un approccio di tipo pluralistico che contempli entrambi i paradigmi.

Oggetto di studio della teoria della complessità sono i sistemi complessi caratterizzati da:

- numerosi elementi interagenti,
- non linearità delle interazioni,
- comparsa a livello globale di proprietà emergenti,
- capacità di auto-organizzazione.

Quindi la scienza della complessità studia quei sistemi dinamici composti dall'interazione di molti elementi e affronta comportamenti globali che non sono comprensibili né prevedibili dallo studio delle caratteristiche delle singole parti; nei sistemi complessi si osservano: (1) dinamiche non-lineari e (2) la comparsa di proprietà emergenti.

Per interazioni non-lineari si intende che molte interazioni tra le componenti avvengono contemporaneamente, regolate da funzioni matematiche non-lineari, o più semplicemente si può dire che i vari elementi formano una rete (Fig. 1).



Fig. 1. Rappresentazione di interazioni di tipo lineare (a sinistra) e interazioni di tipo non-lineare (a destra).

La non-linearità delle interazioni è una caratteristica fondamentale che distingue i sistemi complessi da quelli complicati.

“Complicato” deriva dal latino “*cum plicum*” dove *plicum* indica la piega di un foglio, ovvero “piegato insieme”; “complesso” deriva dal latino “*cum plexum*”, dove *plexum* indica l’intreccio, ovvero “intrecciato insieme”. Un problema complicato può essere risolto con approccio analitico, spiegando il problema nelle sue “pieghe”: si considerano le diverse parti, si trova una spiegazione per ciascuna, le si ricompono e si è risolto il problema. Un approccio di questo tipo è possibile perché è caratterizzato da molte variabili e molte relazioni lineari tra di esse. Un problema complesso invece è caratterizzato da molte variabili e moltissime relazioni lineari e non-lineari; è dunque necessario un approccio sistemico: l’intreccio è impossibile da comprendere nelle sue parti; è necessario quindi ragionare in termini di sistema: bisogna considerarlo nel suo insieme come qualcosa di indivisibile guardandolo da un punto di riferimento esterno.

Nei sistemi lineari il tutto è rigorosamente uguale alla somma delle sue parti, le connessioni non danno nessun valore aggiuntivo. Il comportamento di un sistema complesso, invece, oltre che dai singoli elementi è anche in gran parte determinato dalle loro interazioni. Tali sistemi hanno una natura gerarchica, cioè hanno la tendenza a formare strutture a più livelli, di sistemi dentro sistemi. Ognuno di questi forma un tutto rispetto alle sue parti, mentre allo stesso tempo è parte di un tutto più ampio. Ad ogni livello di complessità, i fenomeni osservati mostrano proprietà che non esistono a livello inferiore, sono state perciò definite dal filosofo C. D. Broad “proprietà emergenti”, dalla parola inglese “*emergence*”, nel senso di proprietà che emergono. Scopo della scienza della complessità è dunque lo studio dell’emergere di queste proprietà collettive (De Toni e Comello, 2005).

Infine i sistemi complessi sono auto-organizzati, possiedono cioè l’incredibile caratteristica di non essere stati progettati da nessuno; in questi sistemi si osserva una manifestazione spontanea di ordine anche in presenza delle incessanti forze di dissoluzione descritte dal secondo principio della termodinamica. Kant fu il primo a utilizzare il termine nella *Critica del giudizio*, in un modo molto simile all’accezione contemporanea: egli si riferiva in particolare agli organismi viventi dove le parti esistono una *per* l’altra, nel senso che si sostengono a vicenda in un tutto funzionante ma allo stesso tempo esistono anche *per mezzo* dell’altra, nel senso che si producono a vicenda. La definizione più rigorosa di auto-organizzazione viene fatta risalire al fisico, premio Nobel, Prigogine negli anni Sessanta e ai primi studi sulla cibernetica, quando alcuni scienziati cominciarono a costruire modelli matematici per rappresentare la logica delle reti neurali. Attraverso l’utilizzo di reti binarie

essi scoprirono che dopo un breve periodo di lampeggi casuali, nella maggior parte delle reti apparivano degli schemi ordinati. Questo tipo di modello sembra essere in grado di spiegare anche fenomeni che hanno una tendenza organizzativa come il lampeggiamento sincrono delle lucciole o il ritmo che a volte diventa perfettamente simultaneo quando centinaia di persone applaudono (Capra, 1997). Strutture auto-organizzanti sono presenti ovunque sia nella materia inanimata sia in natura: nel laser tutti i fotoni sono raggruppati in fase a formare un singolo fascio di luce; le molecole d'acqua che si dispongono in modo ordinato a formare un fiocco di neve; la forma esagonale che le celle di convezione assumono in una pentola di brodo bollente (Waldrop, 1995); le reti di interazioni tra le proteine all'interno della cellula (Barabási, 2004).

All'interno della scienza della complessità, esistono diverse discipline quali per esempio la teoria dei sistemi e la teoria degli automi cellulari; quella che si occupa dello studio delle interazioni tra diverse componenti è la *teoria delle reti*.

Questa nuova scienza delle reti ha preso piede solo negli ultimi anni e si distingue dai lavori precedenti sulla teoria dei grafi in tre punti fondamentali (Newman, Barabási e Watts, 2006):

- si focalizza sulle proprietà delle reti del mondo reale;
- considera le reti come qualcosa che evolve secondo una serie di regole;
- si pone come obiettivo quello di studiare la topologia ma anche il comportamento complessivo delle reti.

## **1. Introduzione alla teoria delle reti**

Matematicamente la rete è un grafo in cui degli oggetti (nodi) sono collegati tra loro da legami (*link*) di varia natura. Una volta adottata questa prospettiva si scopre che 'le reti sono ovunque': in qualsiasi momento, noi siamo il risultato ma anche parte integrante di svariate reti, a partire dal livello microscopico fino a reti che avvolgono l'intero globo terrestre.

Il concetto di rete pervade in egual misura i fenomeni sociali, economici, scientifici e tecnologici; si è scoperto che le proprietà delle reti possono notevolmente influenzare la dinamica di molti processi, che vanno dalla diffusione delle epidemie, alla crescita di Internet, ai processi di comunicazione, per citarne solo alcune. Quando queste mappe sono state accostate una all'altra ci si è resi conto che seguono tutte un'impronta comune, inoltre recenti scoperte sembrano mostrare che principi organizzativi incredibilmente semplici governano la struttura e l'evoluzione di tutte le reti che ci circondano. Queste reti infatti, pur apparentemente così diverse tra loro, sono definite 'piccolo-mondo' perché possiedono una elevata connettività: è possibile passare da un nodo all'altro con un numero molto basso di

passaggi e sono caratterizzate dalla presenza di nodi particolari con un numero elevato di *link* rispetto al resto dei nodi, e sono chiamati *hub* (Fig. 2). Gli hub sono speciali: essendo collegati a un numero insolitamente grande di nodi, accorciano tutte le distanze all'interno del sistema. Questi due concetti verranno meglio descritti in seguito, vengono ora presentate alcune delle reti del mondo reale e nella tabella 1 sono riportati alcuni esempi di reti.

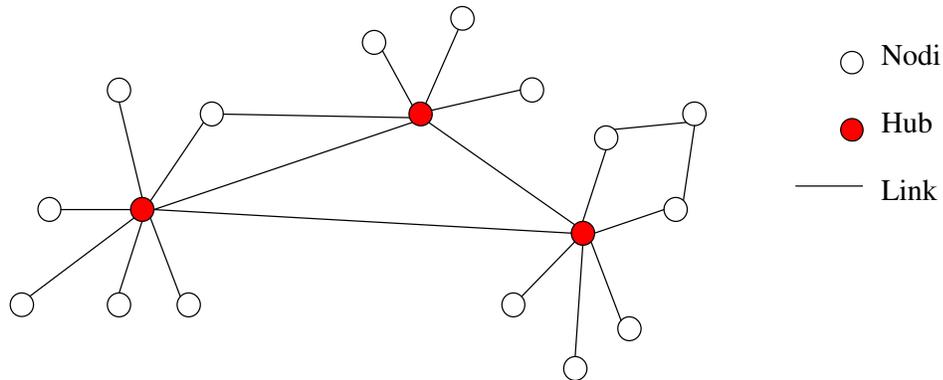


Fig. 2. Rappresentazione delle struttura della maggior parte delle reti reali.

Tab. 1. Alcuni esempi di reti del mondo reale con proprietà di piccolo-mondo.

<b>Reti</b>	<b>Nodi</b>	<b>Connessioni</b>
Reti sociali e di amicizia	Individui	Conoscenza
Rete alimentare	Specie	Relazione preda-predatore
Diffusione delle malattie	Individui (ammalati, portatori sani...)	Meccanismo di infezione
Citazioni scientifiche	Articoli	Citazione e riferimento
Rete degli attori	Attori	Partecipazione allo stesso film
WWW	Pagine web	Collegamento ipertestuale tra pagine
Internet	Router	Linea di comunicazione
Rete metabolica	Substrati e complessi enzimatici	Reazione

## 2. LE RETI DELLA VITA

### 2. 1. Le reti sociali

La sociologia è stata una delle prime discipline a 'pensare a rete': ognuno di noi vive all'interno di un cluster gigante, la rete sociale di tutto il mondo, da cui nessuno è escluso; basta infatti una sola conoscenza per farne parte. Secondo la United Nations Department of Economic and Social Affairs il 12 ottobre 1999 si sono superati i 6 miliardi di persone sul pianeta Terra (Newman, 2000) e nonostante questo enorme numero la struttura delle reti sociali è tale per cui la sensazione è di vivere in un mondo-piccolo. E' capitato a chiunque

infatti, almeno una volta di incontrare casualmente una persona, socialmente e/o geograficamente distante, e di scoprire di avere conoscenze in comune. Ciò significa che siamo connessi con chiunque altro attraverso una breve catena di conoscenze che sarebbe lunga in media solo sei passaggi. Ma oggi viviamo in un mondo sempre più globalizzato in cui poste, telefoni, aerei ma soprattutto Internet permettono di mantenere contatti con persone in ogni angolo del pianeta: i gradi di separazione potrebbero anche essere vicini a tre (Barabási, 2004).

### **Amicizia**

Secondo una famosa espressione ognuno di noi sarebbe separato da solo ‘sei strette di mano con il Presidente degli Stati Uniti’. In realtà bisogna precisare che il Presidente non occupa nessuna posizione privilegiata in questo caso, tale espressione potrebbe valere per qualsiasi persona al mondo (Buchanan, 2003), ma provare a collegarsi con un personaggio famoso rende il gioco più interessante. Io, per esempio, posso dire che tramite tre miei amici mi trovo ‘a solo quattro strette di mano’ dall’etologo Konrad Lorenz e ‘a tre strette di mano’ con Nasser, l’ex Presidente della Repubblica Egiziana, e i due evoluzionisti Stephen Jay Gould e Niles Eldredge (Fig. 3).

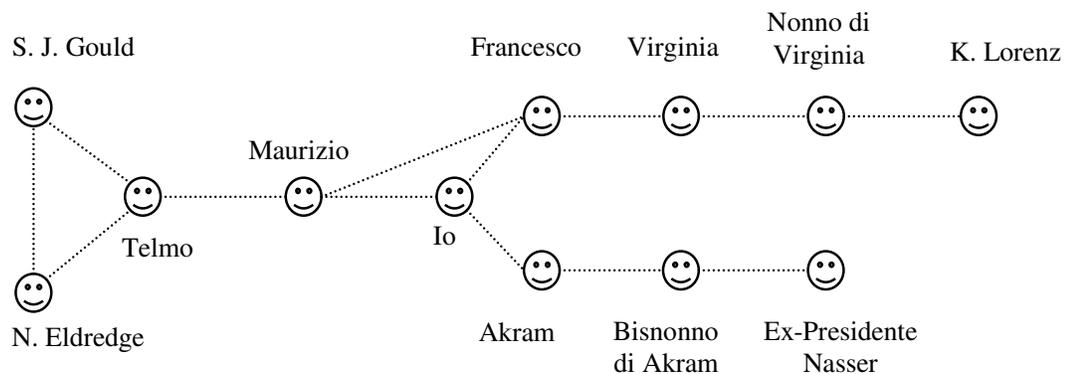


Fig. 3. Rappresentazione delle ‘4 strette di mano’ che mi separano da K. Lorenz e delle ‘3 strette di mano’ che mi separano dall’ex-Presidente della Repubblica Egiziana Nasser e dagli evoluzionisti S. J. Gould e N. Eldredge.

### **La rete delle collaborazioni**

La rete degli attori può essere considerata una buona rappresentazione di una rete sociale. La dimostrazione che anche questa è un piccolo mondo è accessibile a tutti; sul sito [www.cs.virginia.edu/oracle](http://www.cs.virginia.edu/oracle) è disponibile l’*oracolo di Kevin Bacon* ideato, per divertimento, da due laureandi in informatica. Consultando l’Internet Movie Database ([www.imdb.com](http://www.imdb.com)) l’*oracolo* è in grado di calcolare i gradi di separazione tra un qualunque attore e Kevin Bacon; due attori sono connessi tra loro se hanno girato un film insieme; digitando il nome di un

qualsiasi attore l'oracolo risponde dando il numero dei gradi di separazione e rivela anche la catena di film e attori che li collegano. Il risultato, a dir poco eccezionale, è che ogni attore è connesso a Kevin Bacon in non più di due o tre passaggi. Bisogna precisare che Kevin Bacon non occupa nessuna posizione privilegiata; gli stessi risultati si ottengono usando il nome di un qualsiasi altro attore, (è possibile verificarlo nella variante 'Star Links', disponibile sullo stesso sito web), questo perché la rete degli attori è una rete piccolo-mondo.

Il corrispettivo nel mondo scientifico all'oracolo di Kevin Bacon è il cosiddetto "numero di Erdős". Paul Erdős (1913-1996), è il matematico ungherese, padre della teoria dei grafi insieme a Alfréd Rényi; famoso per il suo stile di vita 'vagabondo', è stato tra i più prolifici ed eccentrici scienziati della storia: pubblicò oltre 1500 articoli con 507 matematici (Hoffman, 1999). Chiunque abbia firmato un articolo con lui ha un numero di Erdős pari a uno; chiunque ne abbia firmato uno con un coautore di Erdős ha un numero di Erdős pari a due, e così via. Gran parte dei matematici ha numeri di Erdős molto bassi, i valori sono compresi tra due e cinque gradi di separazione. Ma possono essere collegati a lui anche economisti, fisici, informatici, ecc. Ora è disponibile una pagina web (<http://www.oakland.edu/enp/>) in cui è possibile, per chiunque abbia delle pubblicazioni al suo attivo, calcolare il proprio numero di Erdős. Questo numero dimostra che la comunità scientifica forma una rete altamente interconnessa e valori così bassi indicano che la rete della scienza è davvero un piccolo mondo (De Castro e Grossman, 1999). Infine il numero di Erdős rende l'idea della rete di relazioni che esiste all'interno del mondo scientifico, infatti essere coautori significa avere un legame di conoscenza; anch'essa quindi non è altro che un formato ridotto della rete sociale complessiva.

### ***La diffusione delle epidemie***

Vari risultati ottenuti a partire dallo studio della diffusione dei virus informatici descrivono, con alcune modificazioni, anche il diffondersi di idee, di innovazioni, di nuovi prodotti e anche delle malattie.

Internet è una rete di connessioni fisiche (fili e cavi) fra computer. Satorass e Vespignani (2001) hanno esaminato la diffusione dei virus nella rete di Internet: una volta infettato un *hub* questo può trasmettere il virus a tutti i computer a cui è connesso. I nodi altamente connessi, quindi, sono per il virus informatico un potentissimo mezzo di diffusione e persistenza.

Molte delle malattie sessualmente trasmissibili, inclusa l'AIDS, si diffondono attraverso la rete dei rapporti sessuali. Da uno studio condotto in Svezia nel 1996 su circa 3000 persone

scelte a caso è risultato che un numero molto ristretto di individui ha un'elevata percentuale di rapporti sessuali rispetto al resto della comunità.

Satorass e Vespignani (2002) hanno dimostrato che una immunizzazione uniforme e casuale degli individui non porta all'eradicazione della malattia. Per combatterla efficacemente bisognerebbe concentrarsi sui connettori (*hub*), cioè quegli individui che hanno una maggior probabilità di trasmettere l'infezione: curare pochi individui molto particolari potrebbe essere la chiave per mantenere sotto controllo una malattia o sconfiggerla (Buchanan, 2003).

## **2. 2. Le reti ecologiche**

Le specie negli ecosistemi naturali sono organizzate in reti complesse: ogni specie costituisce un nodo ed è legata ad un'altra se esiste tra esse un'interazione trofica come per esempio la relazione preda-predatore. Camacho *et al.* (2002) hanno studiato la struttura delle reti alimentari di vari ambienti e hanno scoperto che sono caratterizzate dalla stessa topologia: pochi predatori cacciano un numero elevato di prede. Questi e altri risultati hanno un'importanza fondamentale nella gestione dei problemi ambientali: l'eliminazione di specie altamente connesse (per esempio le specie guida) può portare al collasso dell'ecosistema, ma dall'altro lato la protezione di queste poche specie può contribuire alla salvaguardia dell'intero ecosistema.

## **2. 3. Le reti cellulari**

Diversi studi sottolineano l'esistenza e la complessità delle reti all'interno della cellula. Esistono infatti la rete di regolazione genica in cui i nodi sono i geni e le proteine connessi tra loro da interazioni biochimiche; le reti metaboliche in cui i nodi sono le singole molecole connessi e i link rappresentano le reazioni chimiche; le reti di interazione delle proteine dove due proteine sono connesse se partecipano ad uno stesso complesso proteico.

### ***Le reti di regolazione genica***

In una rete di regolazione genica i nodi sono i geni e i link rappresentano l'effetto (attivazione o la repressione) da parte del prodotto di un gene sull'attività di un altro.

Il gene p53 codifica per la proteina P53, un oncosoppressore, che va a legarsi a specifiche sequenze di DNA. La rete p53 è normalmente disattiva ma viene attivata quando le cellule sono stressate o danneggiate e induce l'espressione dei geni *downstream*, quali per esempio quelli che inducono l'arresto della crescita, la riparazione del DNA e l'apoptosi. Sebbene sia ormai evidente il suo ruolo di *hub* all'interno della rete genica, molte domande sono ancora

senza risposta. Infatti si è compreso che non è sufficiente studiare la singola proteina per sconfiggere i tumori ma bisogna considerarla all'interno della sua complessa rete di interazione e segnalazione cellulare (Vogelstain, 2000; Bourdon, 2003).

### ***Le reti di interazione tra proteine***

Nelle reti di interazione proteica due proteine sono collegate da un link se interagiscono fisicamente o funzionalmente tra loro. Jeong *et al.* (2001) hanno mostrato che poche proteine-*hub* altamente connesse hanno un ruolo centrale nell'interazione con le proteine meno connesse; hanno evidenziato inoltre l'importanza che queste proteine hanno per la sopravvivenza nella cellula di lievito. Risultati simili sono stati riportati anche per le reti di interazione di *Caenorhabditis elegans* e *Drosophila melanogaster*.

***Le reti metaboliche*** sono l'oggetto di studio di questo lavoro e verranno perciò trattate in modo più approfondito in seguito, dopo aver presentato in modo dettagliato e anche dal punto di vista matematico i vari modelli di rete fino ad ora proposti.

In tutte queste reti il ruolo chiave è giocato da quei nodi che fanno parte contemporaneamente di più di un gruppo o *cluster*. Sono gli attori che hanno recitato in film di vari generi (l'attore più connesso risulta infatti essere Rod Steiger e non Kevin Bacon), sono le persone capaci di muoversi in ambienti diversi a diversi livelli sociali (l'amico che incontra sempre qualcuno che conosce in qualsiasi posto vada), sono gli scienziati che affrontano problemi di vari campi del sapere (i 'Leonardo da Vinci' che si occupano di arti e scienze).

Lo studio dei sistemi reali attraverso la teoria delle reti si è rivelato un approccio efficace per la loro comprensione: Albert-László Barabási, dell'University of Notre Dame, Indiana, USA è uno dei protagonisti della rivoluzione delle reti e sostiene che essa costituisce un nuovo paradigma per indagare la multiforme varietà del mondo che ci circonda.

## **3. CENNI STORICI: dalla teoria dei grafi alla nuova scienza delle reti**

### **3. 1. La teoria dei grafi**

La nascita dell'interesse per le reti è recente sebbene l'idea di rete o grafo si sia originata circa trecento anni fa. Infatti si è soliti attribuire la prima concezione della teoria delle reti a Leonhard Euler che con un breve saggio del 1736 inaugurò la prolifica branca della

matematica conosciuta come *teoria dei grafi*. Oggi la teoria dei grafi è alla base della nostra concezione delle reti.

Leonhard Euler risolse, con una dimostrazione semplice ed elegante, il curioso problema dei ponti di Königsberg. Königsberg è una piccola cittadina non lontana da San Pietroburgo, situata sulle rive del fiume Pregel, in cui furono costruiti sette ponti per connettere un'isolotto chiuso tra i due rami del fiume. Il quesito era stato posto in questi termini: "È possibile compiere una passeggiata lungo i sette ponti senza mai attraversare lo stesso due volte?"

La grande intuizione di Euler fu di rappresentare il problema come un grafo: sostituì ognuna delle quattro zone di terra con dei nodi e ogni ponte con un link, ottenendo così un grafo con quattro nodi e sette link (Fig. 4).

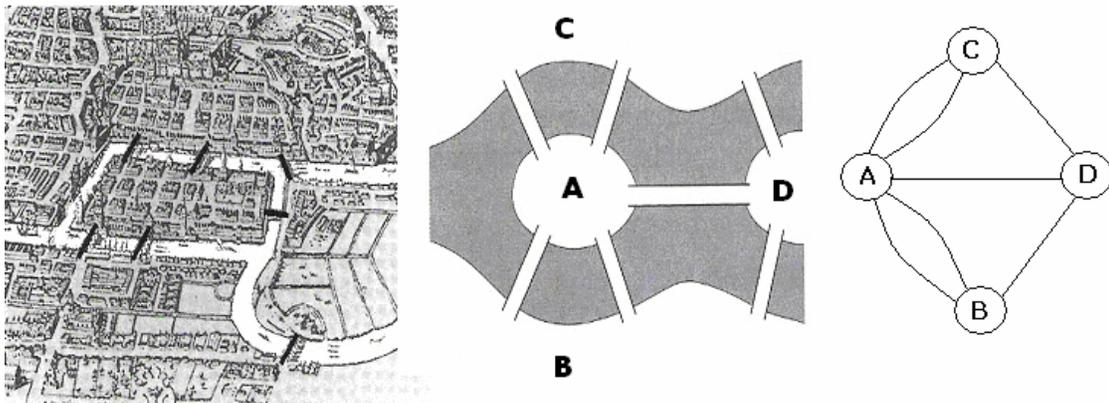


Fig. 4. Rappresentazione del 'problema dei ponti di Königsberg': a sinistra una mappa del XVIII secolo della città in cui sono evidenziati i ponti; al centro un disegno semplificato; a destra il corrispettivo grafo con nodi e link (da Newman, Barabási e Watts, 2006).

Euler osservò che i nodi con un numero dispari di link dovevano trovarsi all'inizio o alla fine del cammino e che su un grafo con più di due nodi e con un numero dispari di link (come il grafo di Königsberg) non poteva essere disegnato un percorso con le caratteristiche richieste. Il problema fu risolto nel 1875 costruendo un nuovo ponte tra B e C, in questo modo restavano solo due nodi con un numero di link dispari: il nodo di partenza e l'ultimo.

Nei trecento anni successivi a Euler, la teoria dei grafi è diventata un campo della matematica vero e proprio, ampiamente sviluppato. Essa definisce un grafo come un insieme di elementi discreti (i nodi), e un insieme di connessioni (*link*) che collegano i vari elementi.

La potenzialità di questa teoria sta nella sua capacità di descrivere molti dei sistemi reali dei più diversi campi, quali l'ingegneria, la sociologia, la biologia, l'informatica. L'importanza della dimostrazione di Euler fu dunque che i grafi possono essere utilizzati come modello per rappresentare la realtà. Infine vedremo che dallo studio delle reti, dalle loro evoluzioni e dalle

loro proprietà è possibile comprendere fenomeni per molto tempo ritenuti troppo complessi o complicati per poter essere spiegati.

Dopo Euler la teoria dei grafi ebbe una vera e propria esplosione, grazie al contributo di importanti matematici come Cauchy, Hamilton, Cayley, Kirchhoff e Polya. Essi scoprirono quasi tutto ciò che oggi sappiamo sulle proprietà dei grafi regolari. Passarono circa due secoli prima che gli scienziati si ponessero la domanda fondamentale: come si originano le reti reali? Cioè, quali sono le leggi fondamentali che fanno evolvere le reti?

I due matematici e amici Paul Erdős e Alfréd Rényi nel 1959 proposero un elegante modello matematico che sembrava potesse spiegare i fenomeni complessi e complicati, così come erano definite le reti reali, soprattutto le reti sociali. Nella loro teoria dei grafi casuali, i due matematici ritengono che le reti reali non sono affatto regolari, ma estremamente complicate, equiparando la complessità alla casualità. In un grafo casuale la probabilità di connessione tra due nodi è uguale per qualsiasi coppia di nodi della rete e si distribuisce in modo casuale. In questo modello basta un numero minimo di connessioni casuali per legare la rete in un insieme quasi completamente interconnesso. In una rete casuale, dunque, è possibile raggiungere qualsiasi altro nodo della rete in pochi passaggi. Sebbene questa proprietà, che venne più tardi definita ‘piccolo-mondo’, sia in grado di spiegare come oltre 6 miliardi di persone siano separati in media da solo 6 passaggi, il modello casuale non coglie molte delle altre caratteristiche delle reti sociali. Se nella rete sociale le conoscenze seguissero una distribuzione casuale allora gli individui avrebbero quasi tutti lo stesso numero di conoscenze e si otterrebbe una società altamente democratica. L’esperienza personale di ciascuno e la sociologia ci dimostrano che la nostra società non funziona così, è molto diversa dall’universo casuale di Erdős e Rényi: esistono cerchie di amici dove si conoscono tutti e pochi legami con l’esterno li mettono in comunicazione con il resto del mondo.

Fu il sociologo Mark Granovetter che nel 1973 nel suo articolo dal titolo ‘La forza dei legami deboli’ (Granovetter, 1983) diede una rappresentazione più realistica delle reti sociali: dimostrò che il segreto del piccolo-mondo era racchiuso nell’intensità dei legami tra le persone. I legami più importanti che mantengono la coerenza di una rete sono paradossalmente i legami deboli, cioè sono le conoscenze più superficiali che fungono da “ponti” tra gruppi di amici. Volendo rappresentare questa situazione con l’immagine delle reti apparirebbe come mostrato nella figura 5: ci sono cluster altamente connessi tra di loro (le cerchie di amici) e pochi ‘ponti’ con l’esterno mettono in comunicazione questi gruppi, evitando l’isolamento.

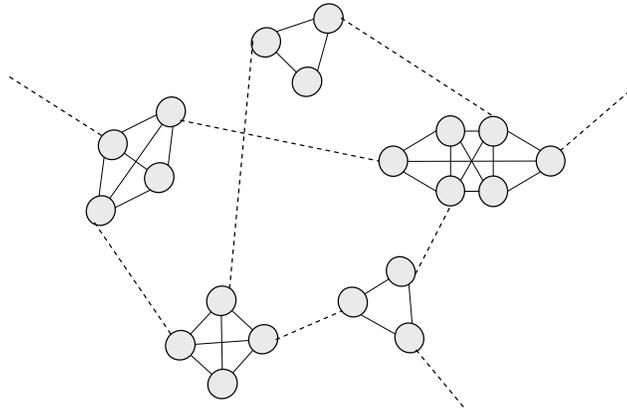


Fig. 5. Rappresentazione della rete sociale con legami forti (linea continua) e legami deboli (linea tratteggiata)

### 3. 1. 1. L'effetto piccolo-mondo

La prima apparizione ufficiale dell'effetto 'piccolo-mondo' risale ad uno scrittore ungherese, molto famoso in patria, Frigyes Karinthy che nel 1929 pubblica un'antologia intitolata 'Ogni cosa è diversa'. Nel racconto 'Catene', l'autore riflette sulla sensazione che "ora, questa terra è più piccola di come non sia mai stata": è il risultato dell'espansione umana, del superamento delle barriere geografiche, delle nuove tecnologie che ci permettono di percorrere l'intero globo in pochi minuti. Racconta che ciascuno di noi è connesso a qualsiasi altra persona sulla terra in non più di 5 passaggi e lo dimostra collegando il protagonista al vincitore del premio Nobel attraverso una catena di conoscenze sorprendentemente corta. Karinthy, dopo aver osservato che collegarsi con un personaggio famoso potrebbe sembrare molto facile, prova a connettere il protagonista con un qualsiasi operaio della Ford e ottiene lo stesso inatteso risultato.

Questa idea di 'piccolo-mondo' è stata resa celebre da alcune prove fornite dal sociologo americano Stanley Milgram (1933-1984). Egli condusse un primo esperimento nel 1967 al fine di studiare la rete di relazioni interpersonali che collega le persone di una comunità. Milgram chiese a un certo numero di persone del Nebraska e del Kansas, scelte casualmente, di inviare delle lettere a un agente di cambio di Boston, di cui era noto solo il nome, senza dare loro l'indirizzo. Le persone che hanno preso parte all'esperimento dovevano recapitare le lettere attraverso la loro rete di conoscenti. Il risultato fu che il numero medio di intermediari necessari era di 5,5 passaggi: una cifra incredibilmente simile a quella suggerita da Karinthy, che venne presto arrotondata a 6.

In realtà i risultati di questo esperimento furono messi in discussione per il loro rigore poco scientifico: nella valutazione dei risultati il numero dei gradi di separazione sarebbe stato

sottovalutato, infatti molte lettere non arrivarono mai all'amico di Boston (quarantadue su centosessanta tornarono indietro). Molte delle lettere mai arrivate potevano aver seguito un percorso molto più lungo prima di essere state buttate (Buchanan, 2003), quindi le catene più lunghe sono rimaste escluse dal campione. Altri invece sostengono che i gradi di separazione possano essere stati addirittura sovrastimati: tra due persone esistono molti percorsi, ogni persona ha deciso di far pervenire la lettera attraverso certe persone e non altre, scegliendo un percorso che non è necessariamente quello più breve (Barabási, 2004).

Nel 1969 Milgram realizzò nuovi esperimenti del genere, al fine di trovare conferma o meno alle proprie precedenti conclusioni. Con la collaborazione di Travers considerò nuovi soggetti e condusse un'analisi quantitativa più dettagliata (Travers e Milgram, 1969). Sia questo che altri lavori simili non fecero altro che corroborare i risultati ottenuti nel primo esperimento.

La famosa espressione "i sei gradi di separazione", forse più nota tra i cinefili e a volte erroneamente attribuita a Milgram, fu coniata nel 1990 nella brillante commedia di Broadway di John Guare dal titolo omonimo. Ma fu grazie al film prodotto dallo stesso Guare, che vede tra i protagonisti Will Smith, che l'espressione raggiunse una grande diffusione popolare. Nel film, una delle protagoniste, meditando sull'interconnettività delle persone dice: "...sei gradi di separazione fra noi e tutti gli altri su questo pianeta. (...) un indigeno australiano, uno della Terra del Fuoco, un eschimese. (...). Ognuno di noi è una porta spalancata su altri mondi."

La pellicola in realtà non ha niente a che fare con il tema della straordinaria interconnessione della rete sociale, ma il semplice fatto di essere stata pronunciata in un film ha fatto sì che questa espressione raggiungesse un pubblico più vasto. In ogni caso il tema delle intricate relazioni umane è un argomento trattato in vari film come per esempio 'Magnolia' di P. T. Anderson (2000) e 'Crash' di P. Haggins (2004).

### **3. 2. Le reti piccolo-mondo tra ordine e casualità**

Abbiamo visto come il modello casuale non tenga conto di un'importante caratteristica delle reti sociali e dei grafi regolari, l'esistenza dei *cluster*: in una rete sociale gli amici di una persona hanno una maggiore probabilità di essere anche amici tra di loro rispetto a due persone scelte a caso.

La valutazione matematica del fenomeno è da attribuire a Duncan Watts e Steven Strogatz (1998) che introducono il fattore chiamato *coefficiente di clustering* o coefficiente di aggregazione per misurare quanto è compatto un gruppo di nodi (o di amici). Un coefficiente di clustering prossimo a 1 significa che il gruppo è molto unito, se invece tende a 0 è molto disgregato. Watts e Strogatz propongono quindi un nuovo modello di reti che contempla un

ridotto numero di passaggi tra una qualsiasi coppia di nodi e l'esistenza dei cluster. Nelle loro reti i due ricercatori collegano i nodi in maniera né ordinata né casuale dando origine a reti che si collocano a metà tra i due estremi: partendo da una rete regolare e aggiungendo delle connessioni casuali ('ponti') ottengono un nuovo modello di rete che chiamano 'piccolo-mondo' proprio in riferimento al fenomeno riportato da Milgram. Questo modello quindi prevede l'esistenza di gruppi di nodi altamente connessi al loro interno e allo stesso tempo collegati agli altri *cluster* da legami deboli. Watts e Strogatz verificano il loro modello misurando i due parametri (coefficiente di clustering e distanza media tra i nodi) in tre reti assai diverse tra loro: la rete neurale di *Caenorhabditis elegans*, la rete elettrica degli Stati Uniti e la rete delle collaborazioni tra gli attori. Confrontano i valori ottenuti con altre tre reti casuali con le stesse caratteristiche (con lo stesso numero di nodi e di connessioni per nodo): in tutte quante le reti il cammino medio tra i nodi è comparabile con quello di una rete casuale mentre il coefficiente di clustering è significativamente maggiore di quello di una rete casuale, tipico dei grafi regolari. Definiscono perciò questo nuovo tipo di rete con caratteristiche intermedie reti piccolo-mondo (Watts e Strogatz, 1998).

### **3. 3. Gli *hub* e le reti a invarianza di scala**

Albert, Jeong e Barabási (1999) condussero uno studio che permise di ottenere la prima mappa della struttura del web: calcolarono quante pagine web avessero uno, due, tre link e così via. Scoprirono con sorpresa che la distribuzione del numero di link sulle pagine web non segue una curva a campana come previsto dal modello casuale ma piuttosto una legge di potenza. Ciò significa che coesistono molti nodi poco connessi con pochi nodi con molti link, definiti *hub*. Hanno anche calcolato il 'diametro del web', cioè quanti *click* di mouse occorrono per raggiungere due pagine web scelte a caso. Il diametro è risultato essere di circa 19, un valore molto basso se confrontato con il miliardo di pagine presenti in rete.

La grande intuizione di Barabási *et al.* è stata quella di capire quanto sia importante la distribuzione secondo la legge di potenza. Una tale distribuzione indica la presenza nella rete di una maggioranza di nodi con pochi link e di pochi nodi dotati di un numero eccezionalmente alto di link, gli *hub*. Questi pochi *hub* hanno un ruolo fondamentale, infatti impediscono alla rete di frammentarsi. Barabási e Albert definirono le reti con questa particolare distribuzione 'a invarianza di scala': tale nome è dovuto al fatto che non è possibile individuare un nodo caratteristico ma si osserva invece una gerarchia continua di nodi che va dai poco diffusi *hub* agli innumerevoli piccoli nodi.

Gli hub si originano secondo il principio dei 'ricchi sempre più ricchi' cioè un nodo con un gran numero di link ha più probabilità di acquisirne altri. Si spiega così anche la famosa espressione "80/20" di Pareto, applicata all'ambito economico: il denaro tende a concentrarsi nelle mani di pochi individui molto ricchi mentre il resto della popolazione guadagna cifre molto scarse.

Diversi lavori dimostrarono o avevano già dimostrato che molte altre reti seguono questa distribuzione:

- la rete degli attori (Watts e Strogatz, 1998),
- la rete delle collaborazioni e delle citazioni scientifiche (Price, 1965; Newman, 1999),
- la connettività dei router di Internet (Faloutsos *et al.*, 1999),
- la rete delle interazioni all'interno della cellula (Jeong *et al.*, 2000),
- la rete linguistica, in cui i nodi sono le singole parole e i link sono rappresentati da significative co-occorrenze in un testo (cioè due parole sono legate se si trovano vicine o più in generale in una stessa frase) (Dorogovtsev e Mendes, 2001).

Una caratteristica fondamentale dovuta alla topologia delle reti a invarianza di scala è la sorprendente robustezza mostrata in caso di errori che eliminano nodi casualmente e invece il crollo della rete nel caso di un attacco mirato agli hub. Infatti i nodi con meno link, che sono anche più numerosi, non contribuiscono granchè all'integrità della rete, se invece viene eliminato anche un numero minimo di hub la rete collassa.

Esattamente in questo modo hanno risposto le reti alimentari alle simulazioni effettuate da Solè e Montoya (2000): gli ecosistemi possono facilmente sopravvivere all'estinzione di alcune specie a caso; se vengono però eliminate le specie più connesse si ha il crollo dell'ecosistema. Così come un blocco simultaneo di aeroporti con traffico molto limitato passerebbe quasi inosservato, mentre la chiusura di poche ore dell'aeroporto di Chicago comporterebbe lo stravolgimento dell'intero traffico nazionale: se chiudessero contemporaneamente gli aeroporti di Atlanta, Chicago, Los Angeles e New York, pur rimanendo aperti tutti gli altri, il traffico aereo degli Stati Uniti sarebbe completamente bloccato (Barabási, 2004).

Lo stesso effetto è stato osservato in seguito a un 'attacco' mirato alla rete delle interazioni proteiche della cellula del lievito. Jeong *et al.* (2001) hanno testato l'importanza di diverse proteine di *Saccharomyces cerevisiae*. La rete ricostruita è formata da 1870 nodi, rappresentanti le proteine, connessi da 2240 interazioni dirette. E' stato dimostrato che le conseguenze fenotipiche della delezione di un singolo gene dipendono dalla posizione topologica del prodotto proteico nella complessa rete delle interazioni. Infatti, la probabilità di

avere un effetto letale rimuovendo una proteina è correlata con il numero di interazioni della proteina stessa: le proteine molto connesse, che hanno quindi un ruolo centrale nell'architettura della rete, sono tre volte più essenziali di proteine con un minor numero di *link*. Un'importante conseguenza di questa particolare struttura è che mutazioni casuali nel genoma non hanno effetti significativi sulla topologia della rete (tolleranza agli errori) a meno che non venga colpito un gene che codifica per una proteina altamente connessa.

#### 4. I MODELLI MATEMATICI PER LO STUDIO DELLE RETI

##### 4.1. I grafi casuali di Erdős e Rényi (1959)

Paul Erdős e Alfréd Rényi sono considerati i padri della moderna teoria dei grafi casuali. I loro lavori furono tra i più influenti e rilevanti tra tutti i lavori pubblicati sull'argomento. Tra il 1959 e il 1968 i due matematici pubblicarono otto articoli sui grafi casuali, che rappresentano i lavori alla base della ricerca sulle reti.

Un grafo casuale è definito come un insieme di  $N$  nodi connessi da  $n$  archi scelti casualmente tra gli  $N(N-1)/2$  archi possibili; in totale ci saranno  $C^n_{[N(N-1)/2]}$  grafi con  $N$  nodi e  $n$  archi. (Erdős e Rényi, 1959). I grafi casuali possono anche essere definiti tramite un modello binomiale: dati  $N$  nodi in cui la probabilità di formare una coppia è pari a  $p$ , il numero totale di archi è una variabile casuale il cui valore atteso è pari a  $p[N(N-1)/2] = E(n)$ .

Quindi la probabilità di avere un grafo  $G_0$  con  $N$  nodi e  $n$  archi è uguale a

$$P(G_0) = p^n (1-p)^{N(N-1)/2-n} \quad (\text{Albert e Barabási, 2002}).$$

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  $Np$  (Przulj, 2004).

Nell'articolo più importante tra quelli riguardanti i grafi casuali i due matematici trattano dell'evoluzione della struttura dei grafi casuali: mostrano come alcune proprietà di tali grafi emergano non gradualmente ma piuttosto improvvisamente quando viene aggiunto un certo numero di link (Erdős e Rényi, 1960).

Un grafo possiede una certa proprietà  $Q$  se la probabilità di avere questa proprietà, cioè  $P(Q)$ , soddisfa questo limite

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

Osservarono che per molte di queste proprietà esiste una probabilità critica  $p_c(N)$  che indica la probabilità con cui il grafo acquisisce tale proprietà. La probabilità che un grafo con  $N$  nodi e con probabilità di connessione pari a  $p(N)$  abbia una proprietà  $Q$  soddisfa il limite

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

Per studiare l'evoluzione dei grafi casuali hanno analizzato la topologia del grafo al variare del numero di archi  $n$ . Con pochi archi,  $n$  piccolo, il grafo risulta frammentato in sottografi o cluster con un numero di vertici al massimo pari a  $O(\log N)$ ; all'aumentare del numero di archi aumenta anche la probabilità che vengono collegati anche nodi che prima erano rimasti isolati e che i vari sottografi si uniscono fra loro. Se viene aggiunto un numero di archi sufficiente in maniera che ogni nodo ne abbia almeno uno, allora emergerà un cluster gigante. Calcolarono questo valore soglia e trovarono che è pari a  $n=N/2$ ; in questa condizione si dice che avviene una transizione o come la definirono i due matematici "emergenza di una componente gigante" che include un'ampia porzione di tutti i nodi e sarà dunque possibile passare da un nodo all'altro (Barabási, 2004).

Erdős e Rényi furono i primi a studiare la distribuzione dei gradi dei nodi nei grafi casuali.

In un grafo casuale con probabilità di connessione  $p$  il grado  $k_i$  del nodo  $i$  segue una distribuzione binomiale di parametri  $N-1$  e  $p$ :

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

Con  $N$  grande la distribuzione dei gradi dei nodi in un grafo casuale si approssima a una Poissoniana (Albert e Barabási, 2002):

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

Nel punto di massimo di questa funzione tutti i nodi hanno all'incirca un numero medio di link; per valori di  $k$  maggiori o minori la funzione decresce rapidamente a indicare che le deviazioni dal numero medio di link per ogni nodo sono rare.

Il diametro di un grafo è definito come la distanza massima, cioè il numero di link, tra una qualsiasi coppia di nodi. Quindi nel caso di un grafo con pochi link (i.e. formato da tanti cluster isolati,  $n$  basso) il diametro sarebbe infinito; ma in un grafo casuale con  $p$  non troppo bassa e  $N$  elevato il diametro risulta piccolo (Albert e Barabási, 2002). Infatti basta che

ciascun nodo abbia almeno un link (cioè  $\langle k \rangle \geq 1$ ) per far sì che si formi un unico cluster gigante in cui tutti i nodi sono connessi.

Albert e Barabási (2002) hanno dimostrato che il diametro di un grafo è proporzionale a  $\ln(N)/\ln(\langle k \rangle)$ ; come osservato in precedenza il diametro rimane basso anche per reti con tanti nodi perché varia in funzione del logaritmo di  $N$ .

Se consideriamo un nodo in un grafo casuale e i suoi nodi adiacenti, la probabilità che due di questi siano collegati è uguale alla probabilità che due nodi qualsiasi scelti a caso siano connessi. Quindi il coefficiente di clustering di un grafo casuale è

$$C_{rand} = p = \langle k \rangle / N$$

#### **4. 2. Le reti piccolo-mondo di Watts e Strogatz (1998)**

Nell'articolo pubblicato nel 1998 da Watts e Strogatz su *Nature* è sottolineato il fatto che molte delle reti reali come le reti neurali, reti sociali o la rete elettrica degli U.S.A., diversamente da come si potrebbe immaginare, si trovano a metà strada tra una rete regolare e una casuale. Questi sistemi, in particolare, presentano un elevato coefficiente di clustering tipico di un grafo regolare e un basso grado di separazione caratteristico dei grafi casuali. Definirono questo particolare tipo di rete *piccolo-mondo*, in analogia con il fenomeno piccolo-mondo descritto da Milgram nel 1967, ora più comunemente conosciuto come 'sei gradi di separazione'.

Nella figura 6 è mostrata la procedura utilizzata da Watts e Strogatz per passare da un grafo regolare a partire da uno casuale ottenendo, in condizioni intermedie (con  $0 < p < 1$ ), un grafo piccolo-mondo.

Watts e Strogatz iniziarono con una rete circolare ordinata: ogni punto è connesso solo con i quattro nodi più vicini; introdussero qualche connessione facendo un collegamento tra due punti presi a caso con probabilità  $p$ . In assenza di connessioni casuali ( $p=0$ ) la rete risulta regolare; all'estremo opposto, cioè quando tutte le connessioni sono casuali, ( $p=1$ ), la rete è casuale. E' per valori intermedi di  $p$  che si ottiene un grafo con caratteristiche particolari, definito poi piccolo-mondo.

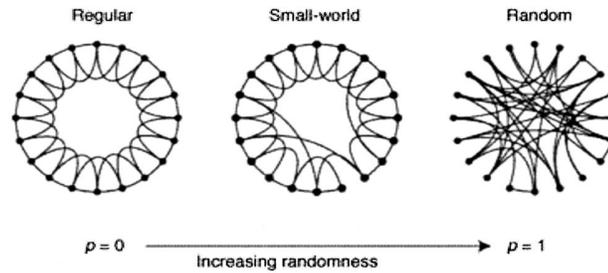


Fig. 6. Procedura utilizzata da Watts e Strogatz per passare da un grafo regolare a partire da uno casuale ottenendo, in condizioni intermedie, una rete piccolo-mondo (da Watts e Strogatz, 1998).

Misurarono quindi le proprietà strutturali di questi grafi:

- il cammino medio tra i nodi,
- il coefficiente di clustering.

Il **cammino medio tra i nodi**,  $L(p)$ , è calcolato come il minimo numero di link necessari per collegare due nodi mediato su tutte le coppie di nodi.

E' interessante notare come varia l'andamento di  $L(p)$  passando da un grafo regolare a uno casuale cioè al variare di  $p$  (Fig. 7); per valori piccoli di  $p$ ,  $L(p)$  varia linearmente con la dimensione del sistema; invece man mano che  $p$  tende a 1 la variazione diventa di tipo logaritmico. Questo brusco cambiamento nell'andamento della curva è dovuto all'aggiunta di un numero minimo di link casuali (*shortcuts*) alla rete inizialmente regolare; così il cammino medio tra i nodi si riduce drasticamente (Watts e Strogatz, 1998).

Watts e Strogatz (1998) utilizzano il **coefficiente di clustering** per misurare il grado di aggregazione della rete.

Il coefficiente di clustering di un nodo,  $C_i$ , è calcolato come rapporto tra il numero di link che realmente esistono nel vertice  $i$  ( $E_i$ ) e il numero di link potenziali, cioè come se il nodo  $i$  fosse collegato a tutti i suoi vicini,  $k_i (k_i-1)/2$ .

$$C_i = 2 E_i / [k_i (k_i-1)]$$

Il coefficiente di clustering di tutta la rete,  $C(p)$ , quindi è dato dalla media di tutti i coefficienti di clustering di ciascun nodo.

Il coefficiente di clustering può assumere valori compresi tra 0 e 1: per valori prossimi allo 0 i nodi sono legati al vertice  $i$  ma non tra di loro; mentre per valori vicini a 1 indicano che i nodi connessi al nodo  $i$  sono anche collegati tra loro.

In una rete casuale la probabilità che due nodi connessi al vertice  $i$  siano anche collegati tra loro è pari alla probabilità che due nodi scelti a caso siano connessi tra loro.

Il coefficiente di clustering è dunque così calcolato

$$C_{rand} = p = \langle k \rangle / N$$

Una rete casuale è poco raggruppata e il cammino medio tra i nodi cresce logicamente con il numero dei nodi.

In una rete regolare il coefficiente di clustering è dato da

$$C = 3(k-2d)/4(k-d)$$

dove  $k$  è il numero di collegamenti di un nodo e  $d$  è la dimensione del grafo; per valori di  $k$  elevati,  $C$  tende a  $3/4$  (Newman, 2000).

Una rete regolare è altamente clusterizzata e il cammino medio tra i nodi cresce linearmente con il numero dei nodi (Fig. 7).

Watts e Strogatz (1998) hanno osservato che esiste un ampio intervallo di  $p$  in cui  $L(p)$  rimane basso come accade in una rete casuale, e  $C(p)$  assume valori molto elevati come accade nelle reti regolari. I due autori hanno dimostrato che esistono reti a metà il casuale e l'ordinato, definite *reti piccolo-mondo*, con caratteristiche del tutto particolari: un basso grado di separazione tra i nodi che allo stesso tempo sono altamente raggruppati.

La straordinaria scoperta è che occorrono pochi link per ridurre drasticamente la separazione media tra i nodi. Questi pochi link formano dei ponti che uniscono nodi situati dalla parte opposta del cerchio facendo così precipitare visibilmente la separazione complessiva tra i nodi ma senza alterare significativamente il coefficiente di clustering.

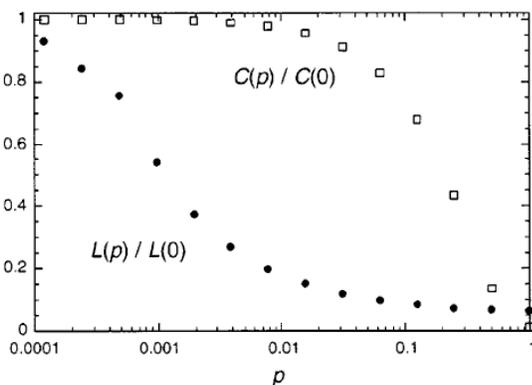


Fig. 7. Andamento del cammino medio tra i nodi,  $L(p)$ , e il coefficiente di clustering,  $C(p)$ , al variare di  $p$  (da Watts e Strogatz, 1998).

Per vedere se le reti piccolo-mondo hanno un riscontro nella realtà Watts e Strogatz hanno misurato il cammino medio tra i nodi,  $L(p)$ , e il coefficiente di clustering,  $C(p)$ , in tre diverse reti del mondo reale: la rete di collaborazioni tra gli attori, la rete elettrica e la rete neurale di *Caenorhabditis elegans*. I valori ottenuti li hanno confrontati con una rete casuale con lo

stesso numero di nodi: tutti e tre i network presentano il fenomeno di piccolo-mondo, cioè  $L(p)$  è uguale o leggermente maggiore a  $L_{random}$  ma  $C(p) \gg C_{random}$  (Watts e Strogatz, 1998). Successivamente oltre a queste reti ne sono state analizzate altre come il World Wide Web, Internet, la rete metabolica di *Escherichia coli* o la rete delle collaborazioni scientifiche. I risultati di tutte le reti analizzate sono riportati in tabella 2: tutte mostrano le proprietà di piccolo-mondo (Albert e Barabási, 2002).

Tab. 2. Caratteristiche topologiche di diverse reti reali analizzate comparate con i valori che una rete con lo stesso numero di nodi assumerebbe se i collegamenti fossero del tutto casuali (da Albert e Barabási, 2002).

Network	Size	$\langle k \rangle$	$\ell$	$\ell_{rand}$	$C$	$C_{rand}$	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	$1.8 \times 10^{-4}$	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	$1.1 \times 10^{-5}$	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	$3 \times 10^{-4}$	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	$5.4 \times 10^{-5}$	Barabási <i>et al.</i> , 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	$5.5 \times 10^{-5}$	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

#### 4. 3. La legge di potenza e le reti a invarianza di scala di Barabási e Albert (1999)

Molti lavori hanno dimostrato che la maggior parte delle reti reali hanno una topologia simile: coesistono pochi nodi con un elevato numero di connessioni (*hub*) e tanti nodi con pochi link; una tale distribuzione si dice che segue una legge di potenza.

Secondo la legge di potenza la probabilità  $P(k)$  che un vertice della rete interagisca con  $k$  altri vertici decade con questo andamento:

$$P(k) \sim k^{-\gamma}$$

L'esponente di grado della distribuzione  $\gamma$ , indica che ogni volta che il numero di connessioni raddoppia, il numero di nodi con quel numero di connessioni diminuisce di  $2^\gamma$  volte (Fig. 8).

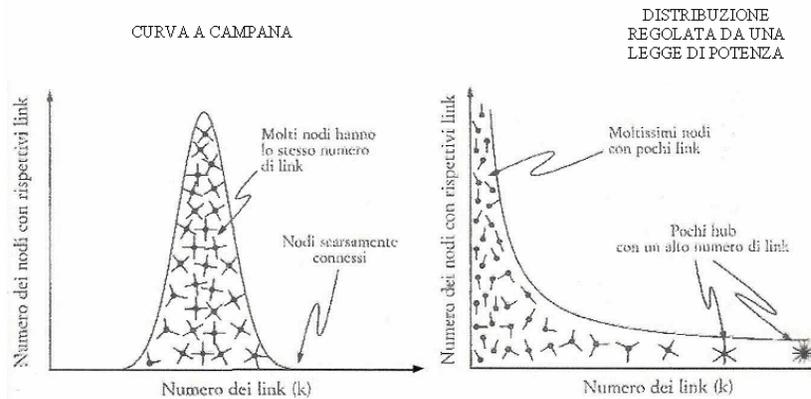


Fig. 8. Distribuzione Normale e distribuzione regolata da una legge di potenza (da Barabási, 2004).

Ma i modelli di reti proposti in precedenza, quello casuale di Erdős (1959) e il piccolo-mondo di Watts e Strogatz (1998), come abbiamo già visto, non mostrano questo tipo di distribuzione.

Questo tipo di distribuzione differisce da quella più comune e più conosciuta, nota come Normale o curva a campana, tipica delle reti casuali; in queste ultime infatti è possibile identificare un nodo caratteristico dato che la maggior parte dei nodi ha lo stesso numero di link e sono rare le deviazioni dalla media. In questo nuovo tipo di reti non è possibile individuare un nodo medio che ben rappresenti tutti i nodi della rete: presentano una struttura gerarchica dei nodi in cui sono rari quelli altamente connessi e numerosi quelli con pochi collegamenti (Barabási, 2004), si può però osservare che un nodo più è connesso più è raro. Barabási e Albert (1999) quindi propongono un nuovo modello che meglio si adatta alle reti reali fino ad allora studiate e le definiscono reti a invarianza di scala o *scale-free network*. Barabási e Albert mostrano come le proprietà comuni a tutte queste reti, possono essere spiegate tramite uno specifico modello dinamico della rete in cui il numero di nodi cresce portando alla generazione di una legge di potenza. Questo nuovo modello differisce da quelli precedenti di Erdős e di Watts e Strogatz nei quali il numero di nodi rimane fisso nel tempo e le connessioni sono casuali, dando origine a una rete egualitaria in cui ogni nodo ha all'incirca lo stesso numero di link; il modello a invarianza di scala è invece in grado di spiegare le reti gerarchiche che mostrano la convivenza di hub e nodi poco connessi.

Sono due i meccanismi, secondo Barabási e Albert, che, in contemporanea, portano alla formazione di una rete a invarianza di scala (Fig. 9):

- la crescita, cioè vengono aggiunti nuovi nodi a ogni intervallo di tempo;
- l'attacco preferenziale, cioè le connessioni non sono casuali ma un nuovo nodo tenderà a connettersi con nodi che hanno già un numero elevato di connessioni.

La probabilità  $\Pi$  che un nuovo nodo,  $j$ , scelga di connettersi al nodo  $i$ , è proporzionale al grado  $k_i$  del nodo

$$\Pi(k_i) = k_i / \sum_j k_j$$



Fig. 9. È mostrato il meccanismo di crescita e attacco preferenziale: all'istante  $t+1$  un nuovo nodo (bianco) è stato aggiunto; i due nuovi link (tratteggiati) sono stati aggiunti in base al principio dell'attacco preferenziale.

Barabási e Albert, per dimostrare che entrambi gli assunti sono necessari per originare una rete a invarianza di scala utilizzano due varianti del modello. Nel modello A si ha solo la crescita della rete ma non c'è attacco preferenziale: un nuovo nodo si connette con uguale probabilità a qualsiasi altro. Il risultato è una rete che cresce ma la distribuzione del grado dei nodi decade più rapidamente della legge di potenza, seguendo un andamento esponenziale: i nodi più vecchi hanno più tempo per collezionare link rispetto agli ultimi arrivati (vantaggio del primo arrivato); con questo modello non si originano dei veri e propri hub. Nel modello B il numero di nodi rimane fisso mentre la probabilità di connessione non è uguale per tutti i nodi; sebbene dopo i primi step la distribuzione sembra seguire una legge di potenza in tempi lunghi si raggiunge una situazione in cui tutti i vertici sono connessi.

### ***Tolleranza ad attacchi ed errori***

Di particolare importanza per qualsiasi genere di rete è la capacità di mantenere un buon grado di connessione anche nel caso in cui alcuni nodi venissero eliminati.

Albert, Jeong e Barabási (2000) hanno verificato l'effetto di attacchi e errori in una rete casuale e in una rete a invarianza di scala. Per *errore* si intende l'eliminazione di nodi in modo casuale; un *attacco* consiste nella rimozione di determinati nodi. La rimozione casuale (*errore*) di una piccola quantità di nodi nelle reti alla Erdős e Rényi, inizialmente non provoca grandi traumi, ma se il numero di nodi eliminati raggiunge un punto critico, il sistema può collassare di colpo frammentandosi in piccole isole sconnesse; il diametro in generale cresce

all'aumentare del numero dei nodi rimossi. Invece nelle reti a invarianza di scala il diametro non aumenta finché non viene eliminato un numero significativamente elevato di nodi: eliminando circa l'80% dei nodi di Internet la rete rimane comunque interconnessa. Responsabile di questo particolare comportamento, definito robustezza topologica, è la presenza degli hub, i rari nodi altamente connessi che tengono insieme la rete: infatti in tali reti la maggior parte dei nodi hanno pochi link, quindi è più probabile che con un attacco casuale vengano colpiti questi che non contribuiscono all'integrità della rete. In ogni caso, neanche la rimozione accidentale di un singolo hub può risultare fatale, perché la gerarchia continua di parecchi grandi hub manterrà l'integrità della rete. Quindi l'alto grado di tolleranza agli errori è legato all'elevata interconnettività, cioè la presenza di rari hub ma anche alla ridondanza: la presenza di vie alternative infatti consente di raggiungere il nodo desiderato anche nel caso in cui alcune delle connessioni che li collegavano in precedenza dovessero essere eliminate. Un *attacco* mirato sulla rete a invarianza di scala, invece, ha un effetto completamente diverso: la rete collassa dopo l'eliminazione di pochi hub. Le reti a invarianza di scala sono dunque robuste verso i guasti ma vulnerabili agli attacchi.

#### 4. 4. Modello a fitness di Bianconi e Barabási (2001)

Il modello proposto da Barabási e Albert implica che i nodi più vecchi avranno per forza un numero di link sempre maggiore dei nodi più recenti. Nel World Wide Web alcune pagine acquisiscono un gran numero di link in brevissimo tempo, fatto che smentisce l'assunto fondamentale del modello a invarianza di scala sopra descritto. Per esempio, un noto caso è il motore di ricerca Google, lanciato nel 1997, era l'ultimo arrivato quando altri motori di ricerca come AltaVista o Inktomi già dominavano il mercato da tempo; in soli tre anni è diventato il motore di ricerca più diffuso. Questo "effetto nuovo arrivato" riscontrato in molte altre reti non è contemplato nel modello a invarianza di scala di Barabási e Albert. Perciò, è stato proposto una variante al modello precedente per spiegare questo fenomeno: il *modello a fitness* di Bianconi e Barabási (2001) dove ogni nodo della rete ha una capacità intrinseca a competere per i link alle spese degli altri nodi. A ogni nodo è assegnato un valore di fitness (cioè l'abilità di guadagnare più *link*),  $\eta_i$ , che non cambia nel tempo; ad ogni intervallo di tempo viene aggiunto un nuovo nodo  $j$  alla rete, con fitness  $\eta_j$ ; ogni nuovo nodo si connette con  $m$  archi ai nodi già presenti nella rete con una probabilità di connettersi al nodo  $i$  che è proporzionale al grado e alla fitness del nodo  $i$ :

$$\Pi_i = \eta_i k_i / \sum_j \eta_j k_j$$

Tale modificazione dell'assunto dell'attacco preferenziale permette anche a un nodo relativamente giovane di acquisire un elevato numero di collegamenti, dato un valore di fitness elevato.

## 5. LE RETI METABOLICHE

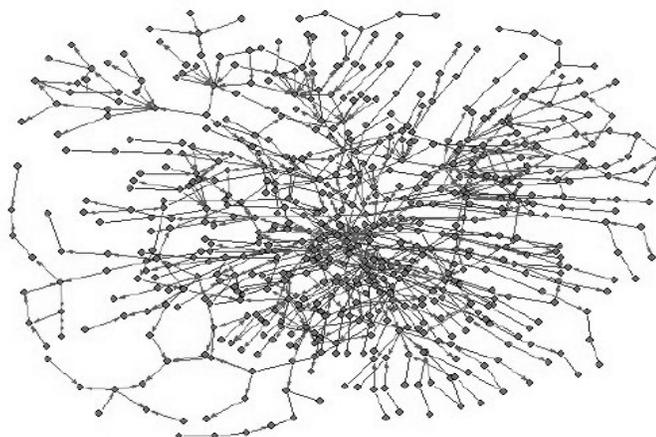


Fig. 10. Rete metabolica di *E. coli* (da Ma e Zeng, 2003)

Nelle reti metaboliche (Fig. 10) i nodi sono le sostanze chimiche, come per esempio l'acqua, l'ATP, il diossido di carbonio, oppure molecole più complesse fatte di dozzine di atomi; i link sono le reazioni biochimiche che li collegano. Così come nelle altre reti descritte in precedenza, ci sono nodi altamente connessi: sono le molecole che hanno ruolo chiave nel metabolismo, partecipano a quasi tutte le reazioni metaboliche, e sono per esempio il piruvato, che costituisce uno dei composti intermedi della glicolisi; il NAD che è un'importante coenzima coinvolto in molteplici reazioni di ossidoriduzione; il CoA è un coenzima che entra nelle vie metaboliche dei carboidrati, degli amminoacidi, degli acidi grassi, dei composti steroidei, dei corpi chetonici e partecipa anche ad altre reazioni di acetilazione.

Il primo lavoro condotto sulle reti metaboliche è stato quello di Jeong *et al.* (2000) che prende in considerazione le reti di 43 organismi rappresentanti tutti e tre i domini della vita. Sono state costruite le reti metaboliche a partire dalle reazioni disponibili sul database WIT e sulle reti ottenute hanno poi condotto un'analisi statistica per identificarne la struttura. Nelle reti costruite i nodi rappresentano i substrati, mentre i link costituiscono le reazioni metaboliche, fisicamente costituite da complesso temporaneo formato dai reagenti e dall'enzima. Considerando separatamente il numero di link in entrata (cioè i link che puntano al nodo) e quelli in uscita (cioè i link che partono dal nodo) hanno riportato su un grafico la probabilità

di un nodo di avere  $k$  link ( $k_{in}$  e  $k_{out}$ ) ottenendo la stessa distribuzione per ogni rete, come quelle mostrate in figura 11, con esponenti  $\gamma_{in}$  e  $\gamma_{out}$  compresi tra 2.0 e 2.4. I risultati ottenuti indicano che la probabilità che un dato substrato partecipi a  $k$  reazioni segue una legge di potenza e ciò significa che le reti metaboliche appartengono alla classe delle reti *scale-free*. Inoltre è emerso che le reti metaboliche di archebatteri, eubatteri ed eucarioti mostrano la stessa topologia indipendentemente dall'esistenza di pathway specie-specifici e ciò significa che la struttura base delle reti si è conservata dal punto di vista evolutivo.

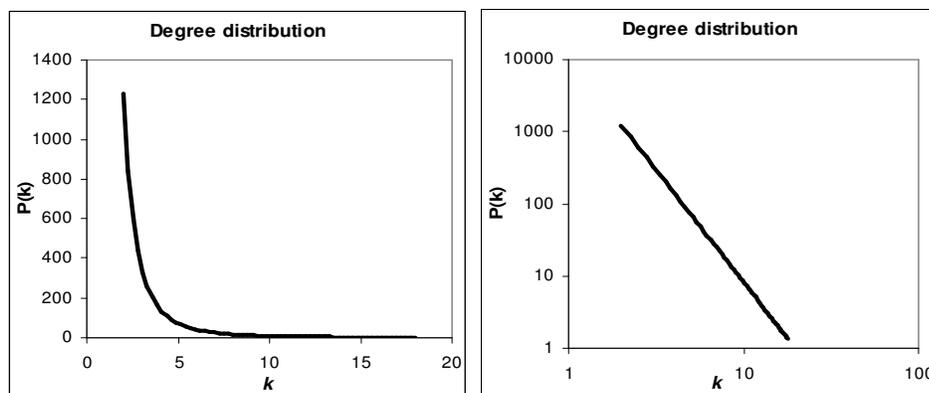


Fig. 11 . Degree distribution  $P(k)$  dei substrati, in scala lineare a sinistra e in scala logaritmica a destra.

Per valutare la proprietà di piccolo-mondo Jeong *et al.* (2000) hanno misurato in ciascuna rete il diametro definito come il percorso più breve esistente tra tutte le coppie di substrati; è risultato che indipendentemente dalla dimensione della rete il diametro rimane costante ed è pari a circa 3 passaggi. Ciò significa che all'aumentare della complessità degli organismi (in riferimento alla dimensione della rete) i nodi sono più interconnessi. Hanno inoltre identificato i metaboliti con il maggior numero di link in entrata e in uscita: le molecole hub risultano essere le stesse per tutti gli organismi e sono acqua, ADP, ortofosfato, ATP, L-glutammato,  $\text{NADP}^+$ , pirofosfato, per citarne solo alcune.

Un'importante conseguenza della distribuzione secondo una legge di potenza è la presenza degli hub; è stata perciò studiata la tolleranza della rete ad attacchi ed errori sulla rete di *E. coli*. Rimuovendo in modo sequenziale i nodi più connessi (simulazione di un attacco mirato agli hub) il diametro della rete cresce rapidamente illustrando l'importante ruolo di questi nodi nel mantenere la rete connessa; eliminando invece i nodi casualmente (per esempio la mutazione casuale in un enzima) la distanza media tra i nodi rimane costante.

Wagner e Fell (2001) hanno successivamente studiato la rete metabolica di *E. coli* costruita in due modi diversi:

- la rete dei substrati, in cui i composti della rete sono collegati se partecipano alla stessa reazione biochimica;

- la rete delle reazioni, in cui i nodi della rete (le reazioni metaboliche) sono connessi se condividono almeno un composto.

Per entrambe queste reti è stato valutato il grado di distribuzione dei nodi, il cammino medio tra due nodi e il coefficiente di clustering. La rete dei substrati risulta essere a invarianza di scala mentre quella delle reazioni è più simile alla rete piccolo-mondi di Watts e Strogaz.

Inoltre hanno analizzato per i due tipi di rete due versioni: una con tutti i nodi e l'altra in cui sono stati rimossi ATP, ADP, NAD, NADP, NADH e NADPH, metaboliti ubiquitari. Tutte e quattro le reti mostrano la proprietà di piccolo-mondo; è possibile cioè passare da un nodo a qualsiasi altro attraverso un numero ridotto di passaggi (circa 3,5 passaggi) oltre ad avere un alto coefficiente di clustering. Hanno inoltre identificato i 13 metaboliti della rete di *E. coli* che hanno il maggior numero di connessioni: sono per la maggior parte molecole coinvolte nel ciclo degli acidi tricarbossilici come acetil coA, ossalacetato, 2-oxoglutarato ecc.

Wagner e Fell propongono inoltre una possibile spiegazione della topologia riscontrata nelle reti metaboliche: i nodi più connessi della rete che si sarebbero originati in base al meccanismo dell'attacco preferenziale (Barabási e Albert, 1999) sarebbero anche quelli più antichi. Spiegano infine come la proprietà piccolo-mondo ottimizzi le funzioni metaboliche: in questo tipo di reti infatti le transizioni si diffondono molto più velocemente e quindi i tempi di risposta alle perturbazioni sono minimizzati.

Ma e Zeng (2003) analizzano la struttura e le proprietà delle reti metaboliche di 80 organismi i cui genomi sono stati completamente sequenziati. Ma e Zeng criticano i risultati ottenuti da Jeong *et al.* (2000): le molecole hub da loro identificate sono comuni a tutti gli organismi perché sono molecole che partecipano a quasi tutte le reazioni, per questo motivo nella loro analisi hanno escluso questi metaboliti ricorrenti. Tra gli hub che hanno identificato, 20 di questi sono comuni alla maggior parte degli organismi. Dall'analisi delle reti analizzate prive dei metaboliti ricorrenti è emerso che mantengono comunque la loro struttura a invarianza di scala. Differenze tra i vari organismi sono state rilevate nel calcolo del diametro (distanza massima tra due nodi) e della distanza media tra due nodi ottenendo risultati in contrasto con quelli di Jeong *et al.* (2000). Eucarioti e archea hanno un cammino medio e un diametro più grandi rispetto ai batteri, a indicare differenze nella struttura delle reti metaboliche; in generale il cammino medio tra due nodi tende a crescere con la dimensione della rete, soprattutto nel caso di reti di piccole dimensioni.

Eom *et al.* (2006) indagano l'organizzazione strutturale delle reti metaboliche di 43 organismi (6 archeobatteri, 32 eubatteri e 5 eucarioti) analizzando i sottografi della rete. È stata identificata l'organizzazione locale (i sottografi) delle reti metaboliche e comparate tra i

diversi organismi. E' risultato che nonostante significative differenze tra i pathway e i substrati dei diversi organismi le proprietà globali e locali sono le stesse. Inoltre l'organizzazione locale delle reti metaboliche mostra caratteristiche che sono peculiari dell'organismo.

Arita (2004) giunge ad una conclusione che contrasta con i lavori precedenti. Costruisce la rete metabolica di *E. coli*, in modo diverso da quelle descritte sopra, collegando i metaboliti in base a legami strutturali: il pathway tra due metaboliti, A e B, è definito come la sequenza di reazioni biochimiche attraverso le quali almeno un atomo di carbonio di A passa a B. Il cammino medio tra i nodi della rete così ottenuta è di circa 8: un valore molto maggiore rispetto ad un grafo casuale con lo stesso numero di nodi e link. Perciò la rete non può essere considerata un piccolo-mondo. Questo metodo di rappresentazione del metabolismo non permette di 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) GLI ORGANISMI ANALIZZATI

In questo lavoro è stata rivolta particolare attenzione alle simbiosi parassitarie, analizzando batteri che provocano patologie nell'uomo e nel caso specifico di *Wolbachia pipientis* che ricopre contemporaneamente un doppio ruolo, di mutualista nei confronti dell'ospite nematode e di parassita per l'uomo nelle filariosi.

Sono state analizzate le reti metaboliche di sei microrganismi:

- *Escherichia coli* e *Salmonella enterica*, due batteri extracellulari, a vita libera;
- quattro specie di batteri endosimbionti appartenenti all'ordine Rickettsiales, *Wolbachia pipientis*, *Ehrlichia chaffeensis*, *Rickettsia typhi* e *R. prowazekii*.

Sono stati scelti batteri che conducono due stili di vita molto diversi, vita libera e intracellulari, per mettere a confronto le loro reti metaboliche e evidenziare eventuali differenze o similitudini.

Si è indagata poi in modo più approfondito la stretta relazione che lega *Wolbachia pipientis* e l'ospite nematode *Brugia malayi*, tra i quali si è instaurata una simbiosi di tipo mutualistico. A tale scopo è stata perciò costruita la rete metabolica di *W. pipientis* endosimbionte intracellulare del nematode filaride *B. malayi* (*wBm*); per lo studio di questa simbiosi sarebbe stato ideale costruire anche la rete metabolica di *B. malayi* ma, dato che il suo genoma è stato da poco sequenziato (è disponibile ora come *draft*) (Ghedin *et al.*, 2007) e che le reazioni metaboliche di questo nematode non sono ancora disponibili nel database PUMA2, è stato scelto il nematode *Caenorhabditis elegans* come organismo modello. Si è pensato infatti che *C. elegans*, ormai molto ben conosciuto potesse rappresentare *B. malayi*. Questo è ovviamente molto approssimativo, e di questa limitazione bisognerà tenerne conto nell'analisi e nella discussione dei risultati, dato che *C. elegans* è un nematode a vita libera mentre *B. malayi* è un nematode parassita; inoltre tra *C. elegans* e *W. pipientis* non è stato documentato alcun tipo di relazione simbiotica.

### 1. L'endosimbiosi

Nel presente studio ci si è concentrati in modo particolare su batteri endosimbionti cioè organismi che vivono all'interno del corpo dell'ospite. Più precisamente *E. coli* e *S. enterica* conducono uno stile di vita extracellulare (cioè nelle cavità dell'ospite) e vengono comunemente definiti 'a vita libera'; le altre quattro specie appartengono all'ordine Rickettsiales e vivono in una condizione intracellulare, cioè all'interno delle cellule dell'ospite.

Vengono definiti simbiotici due o più organismi, appartenenti a specie diverse, che vivono in stretta associazione tra loro; la parola ‘simbiosi’ (dal greco *συν* = insieme e *βίος* = vita), coniata da De Bary nel 1879, comprende relazioni interspecifiche di diverso tipo:

- commensalismo: il commensale trae vantaggio mentre l’ospite non è influenzato;
- mutualismo: ospite e simbiote traggono benefici dalla relazione; si distinguono mutualismo obbligato quando entrambi non sono in grado di vivere senza l’altro e facoltativo quando possono vivere anche separatamente;
- parassitismo: l’ospite è svantaggiato dalla presenza del parassita che manipola o sfrutta le risorse; si distingue parassitismo obbligato o facoltativo.

La categorizzazione dei diversi tipi di associazioni simbiotiche risponde essenzialmente a esigenze di chiarezza e semplicità; infatti, come tutte le definizioni, e soprattutto in biologia, il termine simbiosi comprende una varietà di fenomeni che sfumano uno nell’altro che collegano tra loro le tre principali categorie (Fig. 12). Infatti in molti casi, individuare una netta separazione tra mutualismo, commensalismo e parassitismo è difficile, se non impossibile, in quanto una stessa relazione può assumere aspetti diversi a seconda delle condizioni ambientali.

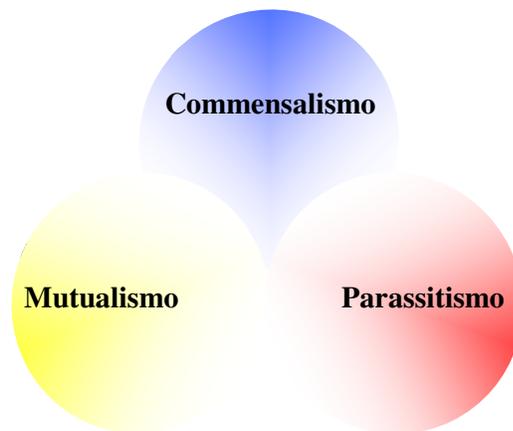


Fig. 12. Rappresentazione del *continuum* che esiste nel concetto di simbiosi.

La distinzione tra batteri parassiti e mutualisti che vivono all’interno delle cellule eucarioti è principalmente effettuata in base all’effetto che il batterio ha sull’ospite. Dal punto di vista batterico entrambi gli stili di vita comportano problemi che vengono risolti in maniera analoga e omologa. Per esempio entrambi gli organismi devono essere in grado di aderire e entrare nella cellula ospite, devono evitare o far fronte al sistema di difesa dell’ospite, devono sopravvivere e moltiplicarsi per infettare nuove cellule o altri ospiti. Le strategie messe in atto da entrambi risultano in somiglianze e differenze caratteristiche dei due stili di vita. In generale molti organismi simbiotici sono difficili da coltivare in laboratorio, altri sono

refrattari alla manipolazione genetica, ma grazie ai nuovi dati disponibili ottenuti dal sequenziamento dei genomi è stato possibile porre le prime basi per comprendere queste associazioni simbiotiche a livello molecolare (Dale e Moran, 2006). Quindi grazie alla disponibilità delle sequenze genomiche di svariati batteri intracellulari obbligati è ora possibile comparare e studiare questi batteri da una prospettiva genomica.

Una simbiosi di tipo mutualistica si riflette in un elevato numero di geni allocati per funzioni di biosintesi; è il caso per esempio di *Buchnera*, un batterio endosimbionte obbligato, che ha instaurato una simbiosi di tipo mutualistica all'interno delle cellule dell'afide: sono stati mantenuti i geni che codificano per la sintesi di amminoacidi essenziali che il batterio fornisce all'ospite; oppure *W. pipientis* mutualista ha mantenuto i geni per la sintesi di alcune vitamine e cofattori che trasferisce all'ospite nematode *Brugia malayi*. La situazione opposta si è verificata nel caso di batteri parassiti come le *Rickettsia* spp. dove solo pochi geni per la biosintesi di amminoacidi sono stati identificati, compensati però da un'elevata frazione di geni che codificano per proteine di trasporto che servono per esempio per importare amminoacidi dalla cellula ospite. Per comunicare con l'organismo ospite, parassiti e mutualisti, hanno evoluto meccanismi per evitare o favorire il riconoscimento: i mutualisti hanno mantenuto geni che codificano per componenti di membrana, proteine e acidi grassi, che servono a farsi riconoscere e rimanere ben ancorati alla membrana della cellula-ospite; all'opposto, i parassiti intracellulari hanno elaborato meccanismi per variare le loro strutture cellulari di superficie in modo da non essere riconosciuti e quindi eliminati dall'ospite. (Tamas *et al.*, 2001).

Un'importante caratteristica dei batteri intracellulari è la riduzione del genoma (o *genome shrinkage*), vengono cioè mantenuti solo quei geni e quei pathway necessari a uno stile di vita simbiotico rendendoli perciò dipendenti dall'ospite. Confrontando il genoma di batteri a vita libera e quello di intracellulari si osserva una significativa differenza, per esempio *E. coli* ha un genoma di circa cinque megabasi che codifica per 5000 geni, mentre il genoma di batteri intracellulari è di circa una megabase e codifica per poco più di 500 geni. I geni mantenuti dagli organismi simbiotici possono essere considerati essenziali se confrontati con il genoma di microrganismi modello e vengono perciò indicati come rappresentativi del 'minimo' genoma batterico (Dale e Moran, 2006). In realtà è sempre più evidente che esistono più soluzioni al problema di mantenere una cellula con un genoma ridotto: è il caso per esempio di *Buchnera* e *Wigglesworthia*, due batteri simbiotici che, sebbene strettamente imparentati, differiscono considerevolmente nel loro set genico: le differenze infatti sono legate a fattori ambientali, al tipo di dieta dell'ospite ma anche alla loro storia evolutiva. I simbiotici infatti

possono originarsi da antenati che hanno un set genico diverso, oppure partendo da un antenato comune divergono in seguito all'adattamento a diverse condizioni (Moran, 2003). Quindi sebbene tutti gli organismi simbiotici siano andati incontro a una riduzione genomica ognuno continua a mantenere caratteristiche uniche e peculiari.

Vengono ora descritti brevemente la biologia e i risultati del sequenziamento dei genomi dei microrganismi oggetto di studio.

## **2. Batteri endosimbionti intracellulari**

### **2.1. Ordine Rickettsiales**

Le specie di batteri appartenenti all'ordine Rickettsiales causano malattie sia nell'uomo che negli animali domestici in tutto il mondo; sono  $\alpha$ -Proteobatteri, parassiti endocellulari obbligati Gram-negativi. Tutti gli organismi dell'ordine Rickettsiales hanno un genoma relativamente piccolo (0.8-1.5 Mb) che si è originato come conseguenza all'adattamento alla vita intracellulare e alla dipendenza dalla cellula ospite per determinate funzioni.

L'ordine Rickettsiales comprende due famiglie:

- la famiglia Anaplasmataceae con i generi *Wolbachia*, *Ehrlichia*, *Neorickettsia* e *Anaplasma*;
- la famiglia Rickettsiaceae con il genere *Rickettsia*.

Le specie appartenenti al genere *Rickettsia* consistono di due gruppi antigenicamente ed ecologicamente ben definiti; tradizionalmente, in base al tipo di malattia che provocano nell'uomo si distinguono:

- il gruppo del tifo (TG), che comprende *R. prowazekii* e *R. typhi*;
- il gruppo della febbre bottonosa o “*spotted fever*” (SFG), comprendente, tra le altre, *R. conorii*;

Altre specie appartenenti al gruppo delle Rickettsiales che provocano malattie nell'uomo sono: *Ehrlichia chaffeensis* che causa la erlichiosi monocitica umana (HME) e infine *Wolbachia pipientis* che ricopre un'importante ruolo nella patogenesi delle filariosi (Foster *et al.*, 2005).

### **2.2. *Wolbachia pipientis***

I circa ottanta anni di ricerche condotte su *W. pipientis* hanno mostrato come questi batteri presentino una elevata diversità molecolare e si trovino in associazione a una gran varietà di specie di ospiti: insetti, ragni, termiti, crostacei terrestri e nematodi filaridi. Ma è solo negli ultimi dieci-quindici anni che è stata rilevata la presenza del batterio in ospiti diversi dagli artropodi.

Hertig e Wolbach nel 1924 osservarono per la prima volta dei batteri intracellulari simili a Rickettsiae negli ovari della zanzara; essi vengono descritti formalmente nel 1936 da Hertig che nomina questa nuova specie di batteri *Wolbachia pipientis* in onore del suo collaboratore Wolbach (Stouthamer, 1999) e in riferimento all'organismo ospite in cui furono ritrovati, la zanzara *Culex pipiens*. I batteri del genere *Wolbachia* sono  $\alpha$ -Proteobatteri, intracellulari obbligati, Gram-negativi appartenenti all'ordine Rickettsiales e della famiglia Anaplasmataceae. *W. pipientis* 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.

Negli anni Cinquanta i lavori di Ghelelovitch e Laven hanno evidenziato come alcuni accoppiamenti tra popolazioni diverse di zanzare portavano alla produzione di una scarsa progenie o del tutto assente. Laven stabilì poi che il fattore responsabile di questa alterazione della riproduzione veniva ereditato tramite il citoplasma e definì questo fenomeno incompatibilità citoplasmatica (CI). Il collegamento tra queste due scoperte, cioè la presenza di batteri intracellulari e la parziale/totale sterilità, avvenne solo nei primi anni Settanta: Yen e Barr stabilirono che il fenomeno dell'incompatibilità citoplasmatica era da associarsi alla presenza di *W. pipientis*; dimostrarono infatti che incroci tra maschi di zanzara infetti con femmine trattate con antibiotici porta alla morte degli embrioni (i.e. incompatibilità unidirezionale).

Negli artropodi sono stati osservati due tipi di incompatibilità citoplasmatica:

- **unidirezionale**, anche se il meccanismo non è ancora del tutto chiaro, è stato proposto il modello modification-rescue secondo il quale la presenza di *W. pipientis* nel maschio porta alla produzione di un fattore (*mod+*) in grado di modificare lo spermatozoo che, successivamente, impedisce lo sviluppo dello zigote, a meno che nell'uovo non sia presente il batterio. Infatti la presenza di *W. pipientis* nell'uovo porterebbe alla produzione del fattore di 'salvataggio' (*res+*) che compenserebbe la modificazione subita dallo spermatozoo. Quindi, come si può osservare dalla tabella, solo un incrocio tra i quattro possibili non ha successo, quello tra l'uovo senza *W. pipientis* (*res-*) e lo spermatozoo modificato (*mod+*). Questo fattore di modificazione non è ancora stato ben identificato, però sembra che esso impedisca alla cromatina di decompattarsi in seguito al legame con alcune proteine istoniche, e ciò impedirebbe anche la prima divisione cellulare (Werren, 1997). Nelle specie di artropodi diploidi si formano embrioni aploidi che vengono eliminati, mentre nelle specie aplodiploidi si formano solo maschi.

		Femmina	
		<i>res+</i>	<i>res-</i>
Maschio	<i>mod+</i>	+	-
	<i>mod-</i>	+	+

- **bidirezionale**, quando in una popolazione di artropodi gli individui sono infettati da diversi ceppi di *W. pipientis*: si verifica incompatibilità riproduttiva nell'incrocio tra uovo e spermatozoo contenenti ceppi diversi del batterio. In questo caso la presenza del fattore *res+* non è in grado di riparare uno spermatozoo *mod+*, è necessario che i batteri siano dello stesso ceppo.

Yen e Barr quindi giunsero ad una conclusione abbastanza sorprendente e cioè che *W. pipientis* alterando la riproduzione di un altro organismo potrebbe costituire una barriera riproduttiva; l'incompatibilità citoplasmatica, infatti, può essere considerata un fenomeno con conseguenze di grossa rilevanza evolutiva. E' stata avanzata l'ipotesi che *W. pipientis* abbia un importante ruolo nella speciazione dell'ospite tramite l'isolamento riproduttivo: se due popolazioni di una stessa specie di artropode sono infettate da due ceppi diversi di *W. pipientis*, allora potrebbe insorgere una barriera (dato che le due popolazioni non danno prole) che potrebbe portare alla separazione in due specie distinte di ospiti; si parla per questo di "speciazione infettiva". Dato tale effetto prodotto da *W. pipientis* negli ospiti artropodi viene coniato un nuovo termine per descrivere questo tipo di simbiosi e cioè 'parassita riproduttivo' identificando quegli organismi che sono in grado di manipolare la riproduzione dell'ospite a proprio vantaggio.

Negli artropodi *W. pipientis* viene trasmessa per via verticale ma ci sono evidenze di trasmissione anche orizzontale; è localizzata nelle cellule degli organi riproduttivi, negli ovari e nei testicoli, nelle cellule dei tubuli malpighiani.

Oltre all'incompatibilità citoplasmatica, *W. pipientis* causa altri effetti negativi sulla riproduzione degli ospiti in modo da garantirsi più successo nella trasmissione:

- **Induzione della partenogenesi:** in popolazioni di imenotteri è stato osservato che *W. pipientis* è in grado di modificare il meccanismo della prima divisione mitotica nello zigote bloccandola all'anafase: nelle uova si forma un nucleo diploide senza che sia avvenuta la fecondazione. Negli organismi aploidi, queste uova si svilupperanno comunque e nasceranno solo individui femmina, spostando perciò il rapporto sessi verso le femmine (Stouthamer *et al.*, 1990).

- **Femminizzazione dei maschi genetici:** sembra che *W. pipientis* sia in grado di interferire durante lo sviluppo di isopodi e lepidotteri a livello della ghiandola androgenica. In questi artropodi la determinazione del sesso è legata alla secrezione di un particolare ormone: *W. pipientis* agisce a livello della ghiandola trasformando i maschi genetici in femmine completamente funzionali (Rousset *et al.*, 1992).

- **Uccisione dei maschi o *male-killing*:** il meccanismo responsabile di questo effetto non è ancora noto ma sembra essere legato a un fattore citoplasmatico che è in grado di uccidere gli embrioni maschi (Hurst *et al.*, 1999).

E' stato osservato che *W. pipientis* si comporta da vero e proprio parassita in *Drosophila melanogaster* (*wMel*): provoca la degenerazione dei tessuti portando alla morte precoce gli adulti infettati.

E' stato infine descritto l'effetto opposto a quanto sopra descritto: *W. pipientis* in questo caso garantisce il successo riproduttivo dell'ospite artropode infettato. Femmine di *Asobara tabida*, una specie di vespa parassitoide di diverse specie di *Drosophila*, se trattate con antibiotici non sono in grado di produrre oociti maturi (Girin e Bouletreau, 1995), mentre i maschi di una specie di coleottero, *Tribolium confusum*, portatori del batterio sono più fecondi.

### ***Wolbachia pipientis* mutualista nelle filarie**

La presenza di *W. pipientis* nei nematodi filaridi è stata riscontrata agli inizi degli anni Settanta da Harada *et al.* che, al microscopio elettronico, osservò corpi tondeggianti negli oociti e negli embrioni di *Dirofilaria immitis*. Solo a metà degli anni Novanta, ulteriori studi di biologia molecolare hanno evidenziato definitivamente che queste strutture trovate all'interno delle cellule del nematode, erano batteri filogeneticamente affini ad altri batteri intracellulari simbiotici, già noti negli artropodi, appartenenti al genere *W. pipientis* (Sironi *et al.*, 1995).

In base ai dati fino ad ora disponibili *W. pipientis* risulta presente solo in due sottofamiglie di nematodi, le Onchocercinae e le Dirofilariinae, della famiglia delle Onchocercidae (Superfamiglia Filaroidea). All'interno di queste due famiglie però risultano specie positive e negative: è stata dimostrata la presenza del batterio nei generi *Onchocerca*, *Dirofilaria*, *Brugia*, *Wuchereria*, *Mansonella*, *Litomosoides*; con prevalenza del 100% invece altre ne sono prive (Casiraghi *et al.*, 2004).

I batteri sono localizzati all'interno delle cellule delle corde laterali in entrambi i sessi, negli oociti, negli embrioni ma non nei tratti riproduttivi dei maschi.

Il nematode e il batterio hanno evoluto una stretta relazione che è sfociata in una simbiosi di tipo mutualistico: *W. pipientis* è fondamentale per la riproduzione e lo sviluppo del nematode, il batterio a sua volta ottiene dall'ospite molecole essenziali per la sua crescita; vari esperimenti condotti per valutare l'effetto di antibiotici nella cura delle filariosi hanno avvalorato l'ipotesi che la simbiosi con *W. pipientis* sia obbligata. Un trattamento con tetracicline ha un effetto inibitorio sulla produzione di microfilarie da parte di adulti oltre a inibire l'embriogenesi: tale terapia infatti sembra provocare una degenerazione di *W. pipientis* e/o una mancata trasmissione del batterio. Ulteriori prove del fatto che *W. pipientis* è fondamentale per la sopravvivenza e la riproduzione della filaria è la presenza del batterio in tutti gli individui della popolazione e la congruenza della sua filogenesi con quella dell'ospite come conseguenza della trasmissione che avviene per via strettamente verticale, cioè dalla madre alla prole (Foster *et al.*, 2005).

Inoltre vari studi hanno dimostrato che una particolare proteina di superficie del batterio, *Wolbachia surface protein* (WSP), nelle filarie viene selezionata negativamente portando a una riduzione della virulenza per poter massimizzare la propria dispersione; la situazione opposta si verifica invece nel caso di *W. pipientis* parassita di *Drosophila melanogaster* (*wMel*): la proteina viene selezionata positivamente permettendo al batterio di variare i propri antigeni in modo da sfuggire al riconoscimento da parte dell'ospite (Jiggins *et al.*, 2002; Baldo *et al.*, 2002).

Data l'assenza di una descrizione formale della specie e data la grande diversità molecolare che *W. pipientis* presenta, si preferisce comprendere in questo nome tutti i rappresentanti riscontrati nei vari ospiti. In base ad analisi di filogenesi molecolare infatti è stato proposto che batteri filogeneticamente affini a *W. pipientis* siano formalmente inclusi in questa specie (Lo *et al.*, 2007).

La filogenesi di *W. pipientis* è stata ricostruita in base ai geni che codificano per 16S rRNA, WSP e *ftsZ* (Foster *et al.*, 2005). Sono stati distinti otto "supergruppi", due dei quali sono ristretti ai nematodi filaridi:

- A e B sono i primi due gruppi identificati negli insetti (Werren *et al.*, 1995),
- C e D nei nematodi filaridi (Bandi *et al.*, 1998),
- E nei collemboli (Vandekerckove *et al.*, 1999),
- F nelle termiti e alcuni nematodi filaridi (Lo *et al.*, 2002),
- G nei ragni (Rowley *et al.*, 2004),
- H in altre termiti, diverse da quelle del gruppo F (Bordestein e Rosengaus, 2005).

Tutte le analisi filogenetiche fino ad ora pubblicate non hanno ancora risolto il problema della radice dell'albero di *W. pipientis*; in assenza di risultati robusti e la continua crescita delle informazioni riguardanti la diversità del genere *Wolbachia* si preferisce mantenere la suddivisione in supergruppi e rappresentare un albero non radicato.

### ***Wolbachia pipientis* endosimbionte di *Brugia malayi* (wBm)**

**Dimensione del genoma:** 1080 kb.

**Numero di geni:** 941.

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

Il genoma di *Wolbachia pipientis* endosimbionte di *Brugia malayi* è costituito da un singolo cromosoma circolare formato da circa 1080 kb e 941 geni. *W. pipientis* ha perso un considerevole numero di geni similmente a quanto osservato anche in *Rickettsia* spp., soprattutto di geni coinvolti nei pathway metabolici.

Grazie alla pubblicazione del genoma di wBm è stato possibile analizzare le sue capacità metaboliche, con particolare attenzione per quelle molecole coinvolte nell'interazione con gli ospiti, la filaria e il mammifero. Sono presenti i pathway completi per la biosintesi di purine e pirimidine; sembra infatti che wBm produca i nucleotidi non solo per un consumo interno ma anche per fornirle all'ospite nematode in caso di necessità, come per esempio durante l'oogenesi e l'embriogenesi. Come accade nelle altre specie appartenenti all'ordine Rickettsiales, le capacità di sintesi di amminoacidi di wBm sono fortemente ridotte: è in grado di sintetizzare un solo amminoacido, il *meso*-DAP. Sono stati rilevati tutti i geni richiesti per la biosintesi di acidi grassi, un gene per la biosintesi di fosfolipidi e quelli che codificano per gli enzimi coinvolti nella biosintesi di riboflavina e FAD: questi coenzimi potrebbero essere trasferiti all'ospite insieme all'eme che sembra essere fondamentale per la riproduzione e lo sviluppo della filaria. È inoltre presente il set di geni per il glutatione che sembra venga fornito all'ospite nematode *B. malayi* sebbene sia in grado di sintetizzarlo, probabilmente è utilizzato contro lo stress ossidativo e per la detossificazione. Infine ha perso circa venti geni coinvolti nella biosintesi di LPS come *E. chaffeensis*; ha perso completamente i pathway, oppure sono presenti solo gli step finali, per la sintesi di cofattori quali biotina, coenzima A, tiamina, acido lipoico, ubiquinone, folato, piridoxina fosfato e NAD (Foster *et al.*, 2005).

### **2. 3. Genere *Rickettsia***

Le due specie oggetto di studio di questo lavoro appartenenti al genere *Rickettsia* sono *R. prowazekii* e *R. typhi* che provocano rispettivamente il tifo epidemico e il tifo murino.

Il ciclo vitale delle specie del genere *Rickettsia* coinvolge sia ospiti vertebrati e invertebrati; in particolare vengono definiti parassiti intracellulari dicensi in quanto svolgono il loro ciclo vitale in due ospiti: un artropode ematofago (zecche, pidocchi e pulci) utilizzato come vettore e il vertebrato, in genere mammifero, che costituisce l'ospite definitivo.

Il meccanismo di inoculazione più frequente è attraverso le feci che il vettore che deposita durante il pasto: se sono presenti lesioni della pelle i batteri possono penetrare nell'organismo ospite, ma la trasmissione può avvenire anche attraverso il morso di pulci o pidocchi. I batteri una volta entrati nel sistema sanguigno infettano principalmente le cellule endoteliali: prima aderiscono alla cellula, inducono la fagocitosi e una volta all'interno si moltiplicano rapidamente. Il meccanismo attraverso il quale il parassita induce la fagocitosi della cellula ospite rimane ancora sconosciuto (McLeod *et al.*, 2004). Le rickettsie del gruppo del tifo inducono la lisi della cellula-ospite da cui escono e vanno poi a infettare altre cellule; quelle del gruppo della “*spotted fever*” attraverso la stimolazione della polimerizzazione delle code di actina che li conduce verso la membrana plasmatica emergono dalla cellula-ospite senza indurre la lisi.

Dall'analisi del genoma delle due specie, *R. typhi* e *R. prowazekii*, è emerso che hanno un elevato grado di similarità e che le minime differenze sono legate soprattutto alla virulenza; infatti, sebbene siano molto simili *R. prowazekii* è altamente patogena per l'uomo mentre *R. typhi* è considerata mediamente patogena (McLeod *et al.*, 2004).

Vengono ora descritte le principali proprietà emerse dall'analisi del genoma che accomunano le specie del genere *Rickettsia*.

I genomi di entrambe le specie appartenenti al gruppo del tifo hanno dimensioni relativamente piccole, di circa 1.1 Mb, caratteristica tipica di batteri intracellulari.

Le specie del genere *Rickettsia* mancano della via glicolitica, infatti non sono in grado di produrre il piruvato e il glicerolo-3P che acquisiscono direttamente dall'ospite (Dunning Hotopp *et al.*, 2006). Una volta entrati nella cellula-ospite, i batteri utilizzano l'ATP già presente nel citosol e solo quando è esaurito cominciano a produrlo da sé attraverso la fosforilazione ossidativa.

Quasi tutti gli amminoacidi devono essere acquisiti dalla cellula-ospite infatti i pathway di sintesi sono assenti o incompleti.

In *Rickettsia* spp. gli enzimi per la biosintesi dei cofattori sono limitati: sono anche assenti i geni per la biosintesi di riboflavina, FAD, tiamina, pantotenato, menachinone, enterochelin e piridossina-fosfato (vitamina B6); è invece ridotto il set di geni richiesto per la biosintesi della biotina e dell'ubiquinone, per questo molti dei cofattori o dei precursori vengono acquisiti

dall'ospite. Inoltre mancano i geni per la sintesi di nicotinate e nicotinamine (i precursori di NAD<sup>+</sup> e NADP<sup>+</sup>), sono però presenti gli enzimi coinvolti nell'interconversione tra NAD<sup>+</sup> e NADP<sup>+</sup>.

Hanno perso la capacità di sintetizzare *de novo* i nucleosidi, sono però presenti i geni che servono per la conversione a nucleosidi difosfato quindi probabilmente i nucleosidi monofosfato vengono importati dalla cellula ospite (Renesto *et al.*, 2005).

Infine sono in grado di sintetizzare il peptidoglicano mentre è molto limitato il set di geni per la sintesi di LPS, inclusi i geni per il lipide A; probabilmente l'assenza di LPS, fattore principale di induzione della virulenza, consente a questi batteri di evitare l'induzione dei meccanismi di difesa dell'ospite (Fuxelius *et al.*, 2007).

### ***Rickettsia prowazekii***

**Dimensione del genoma:** 1111 kb.

**Numero di geni:** 834.

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

Nell'uomo provoca la patologia definita tifo epidemico, infetta le cellule endoteliali del vertebrato e la trasmissione avviene con la puntura del pidocchio (*Pediculus humanus*).

*R. prowazekii* insieme a *R. rickettsii* a causa della loro elevata infettività sono inserite nella lista degli agenti dell'*Antiterrorism and Effective Death Penalty Act* del 1996 (McLeod *et al.*, 2004).

### ***Rickettsia typhi***

**Dimensione del genoma:** 1111 kb.

**Numero di geni:** 838.

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

*R. typhi* è l'agente infettivo del tifo murino e la trasmissione avviene tramite la puntura di pulce del ratto; la malattia ha una diffusione mondiale ma le infezioni sono più frequenti nei climi più caldi tra l'estate e l'autunno mentre la sua distribuzione geografica e stagionale riflette la distribuzione dei vettori (Koliou *et al.*, 2007).

## **2. 4. *Ehrlichia chaffeensis***

**Dimensione del genoma:** 1176 kb.

**Numero di geni:** 1115.

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

I batteri appartenenti alla specie *Ehrlichia chaffeensis* sono patogeni e nell'uomo provocano una malattia sistemica chiamata erlichiosi monocitica (HME). Sono organismi intracellulari obbligati e il loro ciclo vitale coinvolge zecche ematofaghe del genere *Ixodes* come vettore e un ospite definitivo vertebrato. Infetta principalmente monociti e macrofagi in cui si replica all'interno di un comparto delimitato da membrana.

*E. chaffeensis* insieme a *Anaplasma phagocytophilum* e *Neorickettsia sennetsu* sono responsabili nell'uomo diverse malattie che vengono raggruppate sotto il nome di erlichiosi e sono considerate patologie emergenti; dato che in generale i batteri intracellulari sono difficili da coltivare in laboratorio, ancora poco si sa sulla loro biologia. I genomi di questi tre batteri sono stati sequenziati e le informazioni ottenute sono state comparate con i dati già disponibili al fine di migliorare le conoscenze su questi microrganismi in particolare sui fattori di virulenza, sulla patogenesi per lo sviluppo di target per terapie e vaccini.

Dall'analisi del genoma è emerso che *E. chaffeensis* è in grado di sintetizzare tutti i nucleotidi e parte degli amminoacidi (glicina, glutammina, glutammato, aspartato, arginina, lisina).

*E. chaffeensis* a differenza di *W. pipientis* non è in grado di sintetizzare il peptidoglicano; l'ipotesi avanzata è che probabilmente i geni per la biosintesi sono stati persi in modo da consentire l'infezione delle cellule del sistema immunitario dei vertebrati (Dunning Hotopp, 2006) ma come *W. pipientis* è priva di LPS. Inoltre è in grado di sintetizzare la maggior parte di cofattori e vitamine: la biotina è un cofattore essenziale sintetizzato solamente da Anaplasmataceae in grado di infettare vertebrati.

*E. chaffeensis* quindi è dotata della maggior parte dei pathway di sintesi per nucleotidi, vitamine e cofattori, non deve quindi competere con la cellula ospite per questi substrati, inoltre potrebbe fornirli alla cellula ospite.

### **3. Batteri endosimbionti a vita libera**

#### ***Escherichia coli***

**Dimensione del genoma:** 4938 kb.

**Numero di geni:** 4747.

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

*Escherichia coli* è un  $\gamma$ -Proteobatterio, Gram-negativo, appartenente alla famiglia delle Enterobacteriaceae. Fa parte della normale flora batterica dell'intestino inferiore di animali a sangue caldo, dove svolge un importante ruolo nel processo digestivo e può essere quindi considerato un comune commensale. Alcuni ceppi però possono provocare patologie come malattie intestinali, infezioni del tratto urinario, meningite, peritonite, setticemie e polmonite.

In questo lavoro è stato analizzato il ceppo uropatogeno *Escherichia coli* 536 (Uropathogenic *Escherichia coli*, UPEC): è uno degli organismi modello per lo studio delle patologie extraintestinali provocate da *E. coli* ed è la causa più comune delle infezioni del tratto urinario.

### ***Salmonella enterica***

**Dimensione del genoma:** 4755 kb.

**Numero di geni:** 4801.

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

*Salmonella enterica* è un  $\gamma$ -Proteobatterio, Gram-negativo, appartenente alla famiglia delle Enterobacteriaceae.

Il genere *Salmonella* comprende due specie, *S. enterica* e *S. bongori* suddivise in 7 sottospecie e 2400 serovar. E' un microrganismo patogeno per l'uomo e per altri animali; sebbene sia adattato a un ristretto spettro d'ospite, la maggior parte dei serovar presenta un vasto spettro d'ospite.

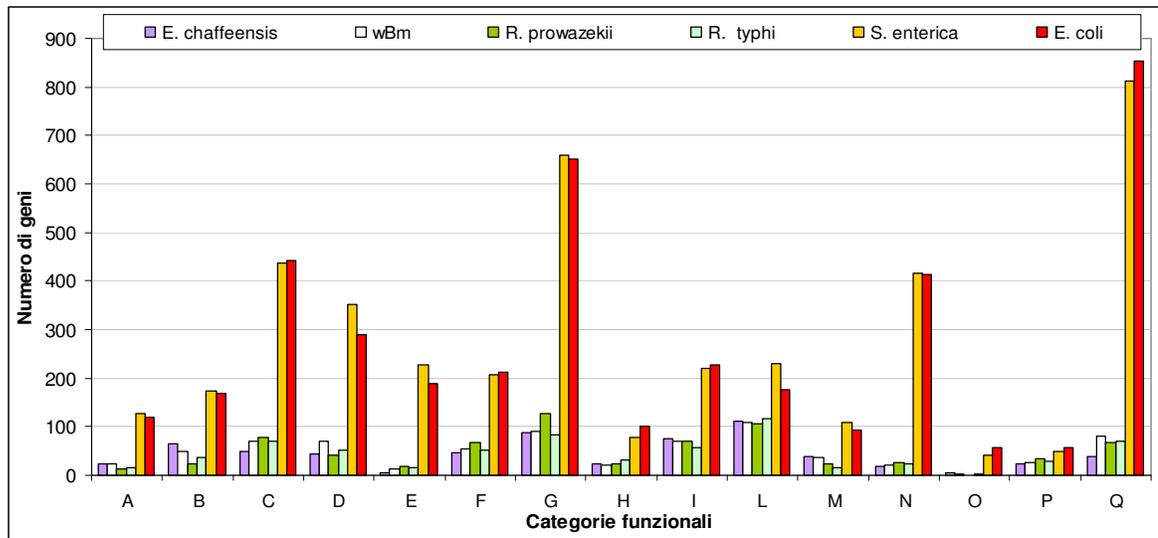
Nell'uomo questi batteri sono gli agenti eziologici delle cosiddette 'salmonellosi' che comprendono due forme clinicamente distinte: la febbre tifoide e la salmonellosi minore; l'infezione si contrae tramite l'ingestione di cibi o bevande contaminate.

La sottospecie presa in esame in questo studio è *S. enterica enterica* serovar Choleraesuis: ha un limitato spettro d'ospiti e infetta occasionalmente anche l'uomo causando gravi aneurismi infettivi.

Nella tabella 3 sono riassunte le caratteristiche genomiche dei batteri considerati nell'analisi e in figura 13 sono riportati, per ciascuna specie, il numero di geni suddivisi nelle varie categorie funzionali (non sono state riportate le categorie particolarmente informative).

Tab. 3. Caratteristiche genomiche dei batteri analizzati, anno di pubblicazione del genoma e stile di vita.

<b>Organismo</b>	<b>Dimensione del genoma (kb)</b>	<b>Numero di geni</b>	<b>Anno di pubblicazione</b>	<b>Stile di vita</b>
<i>E. coli</i>	4938	4747	2006	Vita libera
<i>S. enterica</i>	4755	4801	2005	Vita libera
<i>E. chaffeensis</i>	1176	1115	2006	Intracellulare
<i>W. pipientis</i> (wBm)	1080	941	2005	Intracellulare
<i>R. typhi</i>	1111	838	2004	Intracellulare
<i>R. prowazekii</i>	1111	834	1998	Intracellulare



CATEGORIE FUNZIONALI			
<b>A</b>	Biosintesi degli amminoacidi	<b>I</b>	Protein fate
<b>B</b>	Biosintesi di cofattori e gruppi prostetici	<b>L</b>	Sintesi delle proteine
<b>C</b>	Sviluppo cellulare	<b>M</b>	Purine, pirimidine, nucleosidi e nucleotidi
<b>D</b>	Processi cellulari	<b>N</b>	Funzioni regolatorie
<b>E</b>	Metabolismo centrale intermedio	<b>O</b>	Trasduzione del segnale
<b>F</b>	Metabolismo DNA	<b>P</b>	Trascrizione
<b>G</b>	Metabolismo energetico	<b>Q</b>	Trasporto
<b>H</b>	Acidi grassi		

Fig. 13. Comparazione del contenuto genico (numero di geni) degli organismi analizzati. I geni sono suddivisi nelle loro categorie funzionali, secondo la *TIGR annotation*.

#### 4. Gli ospiti

Vengono ora descritti la biologia dei nematodi e i risultati del sequenziamento dei genomi di *Caenorhabditis elegans* e *Brugia malayi*, ospite di *W. pipientis* (wBm).

##### I nematodi

Le specie di nematodi fino ad ora descritte sono circa 10 mila, ma potrebbero essere ben più di 100 mila quelle esistenti; occupano un'incredibile varietà di nicchie ecologiche: alcuni vivono nei sedimenti marini profondi, altri nei deserti aridi, altri sono batterivori e altri ancora sono parassiti obbligati con vari ospiti intermedi. Più semplicemente si possono individuare tre categorie: nematodi a vita libera, parassiti di piante e parassiti di animali vertebrati e invertebrati.

I nematodi sono vermi cilindrici, di forma affusolata ricoperti da una cuticola traslucida che costituisce un esoscheletro flessibile. Possono essere di dimensioni microscopiche come

*Aphelenchoides bicaudatus* di 0,47 millimetri, ma ci sono specie che possono raggiungere gli 8 metri di lunghezza come la femmina di *Placentonema gigantissima*.

La cuticola è un rivestimento superficiale secreto dall'ipoderma e contenente carboidrati, *mucin-like protein*, glicoproteine e lipidi; in certi casi può presentare particolari strutture annesse come spine o uncini. Sulla superficie dei nematodi parassiti si trovano enzimi che potrebbero inibire la risposta infiammatoria attivata dall'ospite e quindi proteggere il nematode dalle reazioni di difesa dell'ospite (Lee, Cap. VII, 2002).

L'apparato digerente è costituito nella parte anteriore dallo stomodeo che comprende bocca, labbra, cavità boccale e faringe; l'apertura orale è in genere terminale e può presentare organi sensoriali. Lo stomodeo è collegato poi all'intestino che termina nella parte caudale con il proctodeo, un'introflessione ectodermica anale che consiste nell'intestino posteriore; nelle femmine è presente un'apertura anale mentre nei maschi una cloaca.

Molti degli amminoacidi essenziali per i nematodi sono gli stessi che i mammiferi devono ottenere dalla dieta; *C. elegans* e *C. briggsae* richiedono istidina, leucina, isoleucina, lisina, metionina, omocisteina, fenilalanina, treonina, triptofano, valina, arginina. Sono però in grado di sintetizzare alcuni precursori; i venti amminoacidi standard sono utilizzati per sintetizzare la maggior parte delle proteine, amminoacidi minori sono utilizzati per sintetizzare proteine specifiche, altri amminoacidi sono invece utilizzati per funzioni osmotiche (taurina) o come neurotrasmettitori.

L'ammoniaca è il prodotto che viene secreto come risultato del catabolismo degli amminoacidi in particolare dai nematodi parassiti di animali dove è assente un ciclo completo dell'urea.

Sono in grado di sintetizzare *de novo* o a partire dai precursori o da composti intermedi tutti i nucleotidi ma possono anche acquisirli dall'esterno; le specie parassite sembra che non sono in grado di utilizzare la tiamina, citosina, timidina e dTMP.

I nematodi sono in grado di sintetizzare, rimodellare e degradare molte classi di lipidi: elevate quantità di lipidi sono immagazzinate dai nematodi aerobi soprattutto durante gli stadi larvali e gli stadi *non-feeding* J3 di parassiti aerobi e non aerobi, dalle larve dei nematodi a vita libera e nelle uova.

Come gli insetti ma diversamente dai vertebrati e dalle piante, non sono in grado di sintetizzare steroli *de novo* ma devono ottenerli dalla dieta; le specie parassite possiedono livelli relativamente elevati di colesterolo. Gli steroli vengono modificati per sintetizzare gli ormoni steroidei (di cui il colesterolo è la sostanza base per la loro sintesi) che hanno un ruolo importante nei processi di sviluppo e di muta dei nematodi.

La concentrazione dell'acido arachinodico (precursore di molecole segnale) nel nematode *B. malayi* è più alto che nel siero dell'ospite; i nematodi lo assorbono e lo convertono in prostaglandina E2 e prostaciclina, le quali vengono a loro volta secrete nel mezzo e potrebbero essere responsabili della modulazione della risposta immunitaria dell'ospite.

Come tutti gli animali anche i nematodi devono assumere le vitamine attraverso la dieta ma tali richieste nutrizionali non sono state pienamente confermate da esperimenti con terreni di coltura definiti per i nematodi a vita libera, né tanto meno per i nematodi parassiti che sono difficili da coltivare in laboratorio. Studi *in vitro* su *C. briggsae* hanno dimostrato che necessita di cobalamina (o vitamina B<sub>12</sub>), acido folico, pantotenato, tiamina, riboflavina, piridossina, niacinamide e biotina.

*Ascaris suum*, *Brugia malayi* e *B. pahangi* assorbono retinolo (vitamina A) o acido retinoico direttamente dal mezzo di coltura. Il retinolo o i suoi derivati sembrano essere importanti nello sviluppo, riproduzione e metabolismo delle filarie parassite; invece, una quantità eccessiva di retinolo risulta tossica per *C. elegans*: causa una diminuzione della fertilità dopo varie generazioni esposte a livelli elevati. *In vitro* i retinoidi sintetici e naturali inibiscono la motilità e il rilascio delle microfilarie in vari nematodi filaridi (Behm, Cap. X, 2002; Munn, Cap. IX, 2002).

Sono dotati di muscolatura longitudinale che separa l'ipoderma dalla cavità corporea e permette la locomozione con la contrazione. Il sistema nervoso consiste in un ganglio centrale cefalico o di un anello con una serie di nervi cefalici da cui si dipartono le fibre nervose che percorrono in senso longitudinale le corde laterali.

I nematodi sono nella maggior parte delle specie bisessuali e presentano dimorfismo sessuale: per esempio nella specie *Dirofilaria immitis* le femmine possono raggiungere anche i 30 centimetri di lunghezza mentre i maschi sono lunghi circa 12-17 centimetri; alcune specie invece sono partenogenetiche e altre ermafrodite. Nelle femmine sono presenti ovari e uteri doppi, i maschi invece sono provvisti di un unico testicolo che comunica con la cloaca tramite i dotti deferenti (Gibbons, Cap. II, 2002).

### ***Caenorhabditis elegans***

**Dimensione del genoma:** 97 Mb.

**Numero di geni:** 19000 circa.

**Anno di pubblicazione del genoma:** 1998 (The *C. elegans* Sequencing Consortium *et al.*)

Quello di *C. elegans* è il primo genoma di un organismo pluricellulare ad essere stato sequenziato; è formato da 97 Mb con circa 19000 geni (The *C. elegans* Sequencing

Consortium *et al.*, 1998). Altri progetti genoma sono stati completati e altri sono in corso per altri nematodi: è stato completato il genoma di *C. briggsae* e sono in corso i sequenziamenti di genomi di alcuni nematodi parassiti primariamente per la loro importanza in campo medico e agricolo (Hope, Cap. V, 2002).

*C. elegans* è una delle 10 mila specie di nematodi fino ad oggi conosciute; è un piccolo nematode lungo circa un millimetro, appartenente alla famiglia Rhabditidae, è un batterivoro e conduce vita libera nel suolo. Molte delle conoscenze riguardanti l'anatomia, lo sviluppo, la genetica e la genomica dei nematodi sono state acquisite grazie allo studio di *C. elegans*; è considerato infatti un organismo modello per numerosi studi in biologia.

#### **4. 1. Le filarie**

I nematodi filaridi, o filarie, costituiscono un importante gruppo di nematodi parassiti che appartengono all'ordine Spirurida superfamiglia Filarioidea; sono vermi endoparassiti che si localizzano nei tessuti e nelle cavità corporee di vertebrati.

Le filarie sono parassiti eteroxeni, cioè hanno bisogno di uno o più ospiti intermedi per completare il loro ciclo vitale. Vivono all'interno dei tessuti o delle cavità corporee dell'ospite vertebrato, sono ovovivipare o vivipare e rilasciano gli stadi giovanili J1 nel torrente circolatorio, nei vasi linfatici, nei tessuti sottocutanei o nelle cavità corporee; gli individui appena formati sono morfologicamente simili all'adulto e per questo vengono chiamati microfilarie. In seguito alla puntura dell'artropode ematofago che funge da vettore e ospite intermedio le microfilarie raggiungono l'intestino medio. Qui attraverso due mute giungono allo stadio infettivo J3 e si spostano attraverso l'emocele al labium; nel successivo pasto di sangue vengono trasmesse all'ospite definitivo dove, durante la migrazione sottocutanea nel corpo del vertebrato, completano lo stadio J3 e infine quello di adulto. La maturità sessuale viene raggiunta attraverso quattro stadi giovanili (J1, J2, J3 e J4) durante i quali vengono completate quattro mute. L'utilizzo di un ospite intermedio porta alla filaria alcuni vantaggi, come per esempio una maggior dispersione, uno stadio infettivo più longevo, protezione della larva dall'ambiente esterno.

Le filarie provocano nei vertebrati delle patologie dette filariosi che possono assumere forme estreme e gli ospiti definitivi possono essere animali domestici, selvatici e anche l'uomo (Lee, Cap. III, 2002).

## **Filariosi nell'uomo**

Alcuni nematodi filaridi sono patogeni per l'uomo e infettano circa 150 milioni di persone soprattutto nelle zone tropicali e oltre 1 miliardo di persone in più di 90 paesi sono a rischio di infezione (Foster *et al.*, 2005). I nematodi parassiti sono l'agente infettivo delle filariosi cutanee, linfatiche e cardiopolmonare di cane e gatto.

Nell'uomo si distinguono tre principali forme:

- filariosi linfatiche, provocate da *Wuchereria bancrofti* e *Brugia malayi*; causano l'infiammazione delle vie linfatiche in particolare degli arti inferiori (elefantiasi);
- filariosi cutanee: si formano edemi superficiali; le specie responsabili sono *Loa loa* e *Dirofilaria repens*;
- filariosi oculo-cutanee (o oncocercosi), causate da *Onchocerca volvulus*: si sviluppano noduli fibrosi al cui interno si accumulano i parassiti adulti. Le microfilarie possono migrare nella zona oculare e provocare la cecità totale (a questo si deve il nome comune di questa patologia ovvero 'cecità dei fiumi').

## **Il ruolo di *Wolbachia pipientis* nelle filariosi linfatiche**

Le filariosi linfatiche sono le principali infezioni causate da nematodi filaridi nell'uomo; nella maggior parte dei casi sono dovute al nematode *Wuchereria bancrofti*, nei restanti casi sono provocate da *Brugia malayi* e *B. timori*. La patologia nelle filariosi è caratterizzata da infiammazioni acute e croniche; essa è legata soprattutto alle lesioni provocate dalla morte dei vermi adulti che possono risiedere nel sistema linfatico per anni. Con la morte del verme, che può avvenire per via naturale o in seguito a trattamento chemioterapico anti-filarie con dietilcarbamazina (DEC), nel paziente si manifesta una risposta infiammatoria: si registrano elevati livelli di fattore di necrosi tumorale (TNF $\alpha$ ) e interleuchina (IL)-1B, due citochine pro-infiammatorie prodotte dai macrofagi in risposta a stimoli infettivi, accompagnati da episodi febbrili, linfoedema e elefantiasi (Taylor, 2003). Inizialmente era stato ipotizzato che tali manifestazioni fossero dovute a 'tossine' prodotte dai tessuti del parassita; ma poi la precedente ipotesi è stata smentita da Taylor *et al.* (2000) e Brattig *et al.* (2000): esperimenti condotti con estratti di adulti o microfilarie hanno mostrato che questi inducono una potente risposta infiammatoria con produzione di citochine e ossido nitrico; questi risultati hanno suggerito che le molecole che stimolano l'infiammazione hanno caratteristiche simile a endotossine batteriche o al lipopolisaccaride (LPS), sono state perciò definite 'LPS-like'. Ulteriori prove con nematodi naturalmente privi del batterio, come *Acanthocheilonema vitae*, o con linee cellulari infettate da *W. pipientis* e poi trattate con antibiotici infatti non hanno

causato l'attivazione dei macrofagi. Dunque *W. pipientis* sarebbe il responsabile dell'infiammazione nel vertebrato, e quindi anche nell'uomo, in seguito al rilascio di grandi quantità di molecole *LPS-like*, dopo la morte della filaria (Brattig *et al.*, 2000). Risultati simili sono stati ottenuti nello studio sulle oncocercosi, un'altra patologia che colpisce l'uomo, causata dal nematode *Onchocerca volvulus*. In questo caso le microfilarie risiedono nella pelle ma la patologia emerge quando il verme muore; nelle forme più gravi la microfilaria migra nella zona oculare causando una grave infiammazione che può portare a una progressiva perdita della visione e infine alla cecità, da qui il nome comune della malattia 'cecità dei fiumi' (Taylor, 2003). Questi studi hanno mostrato che i vermi vivi rilasciano il batterio e/o i suoi prodotti, (come peptidoglicano inusuale che conterrebbe L-isomeri di alanina e glutammato, *LPS-like*, i dinucleotidi CpG non metilati, la *heat shock protein 60* (HSP 60), i gruppi formil metioninici), che promuovono la risposta infiammatoria nell'uomo o nel mammifero in generale (Bazzocchi *et al.*, 2001). Sembra che a scatenare tale risposta sia, oltre ai composti *LPS-like*, anche una proteina di superficie del batterio, *Wolbachia surface protein* (WSP); è stato dimostrato che WSP viene effettivamente riconosciuta dal sistema immunitario dell'ospite, attiva infatti macrofagi e linfociti del mammifero (Brattig *et al.*, 2004).

Gli studi condotti fin'ora hanno evidenziato l'importante ruolo di *W. pipientis* nelle patologie causate dalle filarie; dunque la presenza del batterio è un aspetto della biologia delle filarie che può essere sfruttato per sviluppare un nuovo approccio per il controllo, l'eradicazione e la cura delle filariosi: sono in corso studi al fine di caratterizzare la natura di queste molecole con attività simile a endotossine e sono in prova terapie con antibiotici. Infatti le strategie precedenti prevedevano o l'eliminazione dell'artropode vettore o l'utilizzo di chemioterapici filaricidi; dietilcarbamazina, albendazole, ivermectina hanno effetto quasi nullo sui vermi adulti e inoltre il trattamento deve essere portato avanti per molto tempo e la possibilità che insorga una resistenza non è da trascurare (Taylor, 2003).

### ***Brugia malayi***

**Dimensione del genoma:** 95 Mb.

**Numero di geni:** 11500 circa.

**Anno di pubblicazione del genoma:** 2007 (Ghedin *et al.*)

*Brugia malayi* (ordine Spirurida, famiglia Filariidae) è un nematode parassita diffuso nel sud-est dell'Asia e dell'Indonesia. Come altri nematodi filaridi si sviluppa attraverso quattro fasi

larvali, interamente all'interno delle specie ospiti, che possono essere artropodi-vettori (*Culex*, *Aedes* e *Anopheles*) e uomo (l'ospite definitivo), dove può vivere anche per decenni.

La recente pubblicazione del genoma di *B. malayi* ha evidenziato che, oltre ad avere un genoma di dimensioni più piccole rispetto a *Caenorhabditis* spp. sequenziati, i set di geni per la sintesi di purine, eme e riboflavina sono quasi del tutto incompleti. *B. malayi* potrebbe acquisire questi metaboliti chiave dall'ospite oppure da *W. pipientis* che ha questi pathway di sintesi completi. Le differenze riscontrate nel contenuto e nell'organizzazione del genoma di *B. malayi* e *C. elegans* hanno sottolineato l'importanza di ottenere ulteriori informazioni che meglio possano rappresentare tutta la diversità del phylum Nematoda; si potranno infatti così individuare i pathway tipici delle specie parassite che potrebbero costituire la chiave per lo sviluppo di farmaci e vaccini per la cura delle filariosi (Ghedin *et al.*, 2007).

### **3) METODI**

#### **1. Costruzione delle reti metaboliche**

##### **1. 1. Il database**

Per la costruzione delle reti metaboliche è stato utilizzato il database PUMA2 disponibile in Internet all'indirizzo <http://compbio.mcs.anl.gov/puma2>, nella sezione "EMP Metabolic Reconstruction" dove, al momento, sono presenti i pathway metabolici di più di duecento organismi.

Il database PUMA2 integra i dati di più di 20 database (tra cui NCBI, PIR, UniProt, KEGG e ENZYME) che riportano informazioni metaboliche, annotazioni delle sequenze, tassonomia e fisiologia, e li assembla con vari strumenti bioinformatici (Maltsev *et al.*, 2006).

##### **1. 2. Costruzione della rete metabolica**

Tramite due programmi scritti in linguaggio Perl in ambiente WIN (Comandatore, unpublished) è stato possibile scaricare le reazioni e quindi costruire le reti metaboliche degli organismi presi in esame.

Il primo programma è stato utilizzato per consultare il database PUMA2 direttamente da Internet: il programma scarica tutte le pagine HTML riguardanti le reazioni metaboliche dell'organismo selezionato e le trascrive in un file TXT. E' stato necessario esaminare e poi correggere manualmente i dati scaricati dal database di ciascun organismo dato che in molti casi uno stesso substrato è risultato essere scritto con nomi diversi, come ad esempio pirophosphate, pyrophosphate e pyrophosphat.

Una volta ottenuta la lista corretta delle reazioni metaboliche per ogni organismo, un secondo programma ha permesso di costruire la rete.

Le reti costruite sono così composte:

- i nodi rappresentano tre tipi di molecole: reagenti, prodotti e centri di reazione; i centri di reazione rappresentano uno o più enzimi che catalizzano la reazione che trasforma i reagenti in prodotti;
- i link sono direzionali: collegano il reagente con il centro di reazione, e poi quest'ultimo al prodotto (Fig. 14).

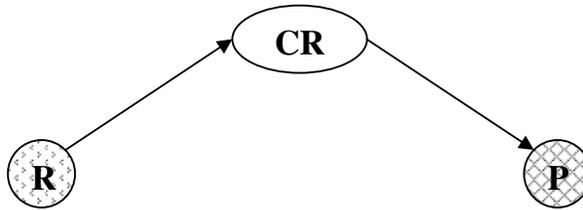


Fig. 14. Rappresentazione di una reazione delle reti metaboliche costruite con un reagente (R), un prodotto (P), un centro di reazione (CR).

Infine per analizzare e rappresentare le reti è stato utilizzato il software Pajek (Batagelj e Mrvar, 1998).

Il file di input richiesto da questo programma (.net) è formato da due parti distinte (Fig. 15):

- nella prima parte, denominata *\*vertices*, è indicato il numero totale dei nodi della rete, successivamente sono elencati tutti i nodi e a ciascun nodo è assegnato un numero;
- nella seconda parte, denominata *\*arcs*, sono elencate coppie di numeri (cioè i nodi) a indicare l'esistenza di un link tra i due nodi; la direzione del link è mostrata dall'ordine di lettura dei numeri: il numero a sinistra indica il nodo di partenza e il numero a destra il nodo a cui punta; il numero 1 a fianco della coppia di numeri indica l'esistenza di un link che unisce i nodi.

```

*vertices 1659
1 <Reazione0<
2 <Reazione1<
3 <Reazione2<
4 <Reazione3<
5 <Reazione4<
...
1103 <ADP<
1104 <orthophosphate<
1105 <sulfate<
1106 <thiosulfate<
1107 <3'-phosphoadenylylsulfate<
1108 <pyrophosphate<
1109 <O(,2)<
1110 <H(,2)O(,2)<
...

*arcs
1105 1 1
1474 1 1
1125 1 1
1111 1 1
1 1105 1
1 1102 1
1 1103 1
1 1104 1
1105 2 1
1474 2 1
1125 2 1
1111 2 1
2 1105 1
2 1102 1
...
  
```

Fig. 15. File di input richiesto dal programma.

## 2. Analisi delle reti metaboliche

Per ogni rete ottenuta, una per ogni organismo, sono stati valutati diversi parametri al fine di dare una descrizione quantitativa della rete:

- Caratteristiche topologiche: numero totale dei nodi, numero dei substrati, numero dei centri di reazione, numero di link;
- Distribuzione dei gradi dei nodi;
- Diametro e distanza media tra i nodi;
- Hub;

## 2. 1. Distribuzione dei gradi dei nodi

Il grado di un nodo è definito come il numero di link posseduto da ciascun nodo. Nelle reti metaboliche analizzate i link sono direzionati, cioè in un nodo si distinguono link in entrata e link in uscita; il link in uscita dal nodo indica che prenderà parte alla reazione come reagente, un link in entrata significa che costituisce il prodotto di una reazione.

Per ciascun nodo è stato calcolato il grado in entrata ( $k_{in}$ ) cioè il numero di link che puntano al nodo e il grado in uscita cioè il numero di link uscenti ( $k_{out}$ ) dal nodo.

È stata quindi calcolata la frequenza per ogni valore di  $k_{in}$  e  $k_{out}$  e riportata su un grafico in scala logaritmica; per verificare l'esistenza di invarianza di scala, i punti sono stati interpolati con la curva detta 'legge di potenza' e sono stati calcolati gli esponenti di grado,  $\gamma_{in}$  e  $\gamma_{out}$  e il valore dell' $R^2$ .

## 2. 2. Diametro e distanza media tra i nodi

Tramite il software Pajek è stato possibile calcolare il diametro (D) e la distanza media (DM) tra i nodi:

D è la distanza massima misurata tra tutte le coppie di nodi;

DM (detta anche cammino medio) è il valore medio calcolato su tutte le distanze tra ogni coppia di nodi.

## 2. 3. Hub

Si definiscono hub i rari nodi altamente connessi, cioè quei nodi con un numero elevato di link, che tengono insieme la rete. Conducendo questa analisi quindi si è voluto individuare quelle molecole che hanno un ruolo chiave nel metabolismo.

In ciascuna rete gli hub sono stati identificati in base alla definizione proposta da Wagner e Fell (2001): un nodo è considerato hub se possiede un numero di link ( $k$ ) maggiore della somma tra la media dei gradi dei nodi ( $k_m$ ) e tre volte la deviazione standard:

$$k > k_m + 3\sigma$$

## 3. Identificazione delle molecole di interazione

Data la struttura delle reti metaboliche con link direzionati si è assunto che le molecole con solo link in entrata (quindi con grado  $k_{out}$  pari a 0) possono essere considerate esclusivamente dei prodotti e quindi, ipoteticamente, secrete dall'organismo. Allo stesso modo le molecole

che hanno solo link in uscita (cioè con grado  $k_{in} = 0$ ) prendono parte alle reazioni solo come reagenti (figura ---) e vengono perciò acquisite dall'esterno (Fig. 16).



Fig. 16. Molecole 'solo reagente' ( $k_{in} = 0$ ) a sinistra e 'solo prodotto' ( $k_{out} = 0$ ) a destra.

Sono stati individuate quindi le molecole 'solo reagenti' e 'solo prodotti' per ciascuna specie. Le liste di molecole ottenute sono state utilizzate per:

- studiare l'interazione ospite – simbiote;
- valutare il tipo di molecole prodotte e acquisite e operare un confronto fra gli organismi;
- costruire due alberi filogenetici, uno per i reagenti e uno per i prodotti.

### 3. 1. Analisi dell'interazione ospite – simbiote

Per studiare le interazioni tra *W. pipientis* e il nematode sono state individuate quelle molecole che i due simbioti si scambiano: sono stati confrontati quindi i prodotti del batterio con i reagenti dell'ospite e poi i reagenti del batterio con i prodotti dell'ospite.

Infine sono state analizzate le molecole di interazione che vengono acquisite e secrete dai microrganismi patogeni includendo nell'analisi anche *W. pipientis* ipotizzando che alcuni dei composti presenti in queste liste possono in qualche modo interagire anche con l'ospite del nematode e cioè il vertebrato (e quindi anche l'uomo).

### 3. 2. Indici di somiglianza

Le liste di molecole delle Rickettsie sono state confrontate con le liste di *E. coli* in modo da individuare quelle molecole che non sono presenti in quest'ultimo e quindi caratteristiche delle quattro specie di Rickettsie. Per valutare la similarità tra queste liste di molecole tra le specie Rickettsie sono stati calcolati gli indici di somiglianza di Sorensen (S) e di Jaccard (J):

$$S = 2c/(a+b)$$

dove  $c$  è il numero di molecole che hanno in comune la specie A e la specie B,  $a$  è il numero di molecole della specie A,  $b$  è il numero di molecole della specie B.

$$J = c/s$$

con  $c$  che rappresenta il numero di molecole che hanno in comune la specie A e la specie B,  $s$  è il numero delle molecole totali tra la specie A e la specie B contando una volta sola le molecole in comune.

### **3. 3. Costruzione dell'albero filogenetico a partire dalle liste di 'solo prodotti' e 'solo reagenti'**

Per tale analisi sono state considerate separatamente le liste delle molecole in entrata e in uscita di tutte le specie. Ogni molecola è stata considerata come se fosse un carattere discreto; è stata quindi valutata la presenza o l'assenza di tale molecola nelle liste di ciascun taxon. Tramite un programma scritto in linguaggio Perl in ambiente WIN (Comandatore, unpublished) i dati sono stati organizzati in due matrici presenza-assenza delle molecole rispetto ai taxa considerati.

Con le matrici presenza-assenza così ottenute sono stati costruiti due alberi filogenetici utilizzando il programma CLIQUE ((c) Copyright 1986-1993 by Joseph Felsenstein and by the University of Washington) disponibile in Internet.

CLIQUE è un programma per la ricostruzione della filogenesi a partire da caratteri discreti (0 e 1) restituendo un albero non radicato; utilizza il metodo di compatibilità cioè trova il set di caratteri più ampio possibile che siano simultaneamente compatibili tra loro.

Il programma si basa sulle seguenti assunzioni:

1. Ogni carattere evolve indipendentemente;
2. Differenti lignaggi evolvono indipendentemente;
3. Lo stato ancestrale è sconosciuto;
4. Ogni carattere ha una piccola probabilità di evolvere molto rapidamente;
5. La probabilità di avere due cambiamenti in un carattere con basso tasso di cambiamento è minore della probabilità che ciò avvenga in un carattere con elevato tasso di cambiamento;
6. Non sappiamo in anticipo quali sono i caratteri con alto o basso tasso di cambiamento.

Per visualizzare gli alberi filogenetici ottenuti è stato utilizzato il programma TreeView (Page, 1996).

#### **4. Costruzione dell'albero filogenetico**

Per chiarire le relazioni filogenetiche tra gli organismi in esame è stato costruito l'albero filogenetico sulla base delle sequenze del 16S rRNA e dei geni *groEL* e *ftsZ*; le sequenze sono state scaricate dalla banca genica disponibile on-line fornita da NCBI (National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/>), sono state corrette con il programma Bioedit 7.0.0 (Hall, 1999) e poi allineate con Clustal X (Thompson *et al.*, 1997). Per l'analisi filogenetica è stato utilizzato MEGA 3.1 (Kumar *et al.*, 2004) e come approccio è stato impiegato il metodo *Neighbor Joining*, un algoritmo che utilizza le distanze esistenti tra ogni coppia di sequenze per costruire gli alberi. Si è applicata la correzione di Kimura.

## 4) RISULTATI E DISCUSSIONE

### 1. Analisi delle reti metaboliche

Sono state analizzate le reti metaboliche di 7 organismi di cui

- 2 specie di batteri a vita libera: *Escherichia coli* e *Salmonella enterica*;
- 4 batteri intracellulari appartenenti all'ordine Rickettsiales: *Ehrlichia chaffeensis*, *Rickettsia typhi* e *R. prowazekii*, organismi patogeni anche per l'uomo, e *Wolbachia pipientis* endosimbionte del nematode *Brugia malayi*;
- *C. elegans*, nematode a vita libera, è stato utilizzato come nematode-modello per l'analisi dell'interazione con *W. pipientis*.

Per ogni organismo è stata costruita e analizzata la rete metabolica e sono stati misurati i seguenti parametri:

- Caratteristiche topologiche: numero totale dei nodi, numero dei substrati, numero dei centri di reazione, numero di link;
- Diametro, cioè la distanza massima misurata tra tutte le coppie di nodi;
- Distanza media tra i nodi, cioè il numero medio dei gradi di separazione;
- Distribuzione dei gradi dei nodi (in entrata e in uscita);
- Identificazione degli hub, cioè i nodi con un elevato numero di link.

### Topologia

Nella tabella 4 e nel grafico in figura 17 sono riassunte le caratteristiche topologiche delle reti degli organismi analizzati.

Tab. 4. Caratteristiche topologiche delle reti metaboliche degli organismi analizzati, dimensione del genoma e numero di geni. N = numero totale dei nodi, S = numero dei substrati, CR = numero di centri di reazione, L = numero di link, D = diametro, DM = distanza media.

Organismo	Genoma (kb)	Geni	N	S	CR	L	D	DM	$\gamma_{in}$	$\gamma_{out}$
<i>E. coli</i>	4938	4747	3263	1175	2088	14063	14	4,38	-3,24	-3,50
<i>S. enterica</i>	4755	4801	3151	1120	2031	13768	13	4,25	-3,91	-3,27
<i>E. chaffeensis</i>	1176	1115	1697	646	1051	8124	12	3,97	-3,10	-2,64
<i>R. typhi</i>	1111	838	1512	624	888	6453	13	4,44	-2,75	-2,81
<i>R. prowazekii</i>	1111	834	1543	668	875	6485	14	4,60	-2,77	-2,83
<i>W. pipientis</i>	1080	941	1658	558	1100	8280	13	3,93	-3,11	-2,83
<i>C. elegans</i>	97000	19000	2101	750	1351	9525	14	4,15	-3,14	-3,00

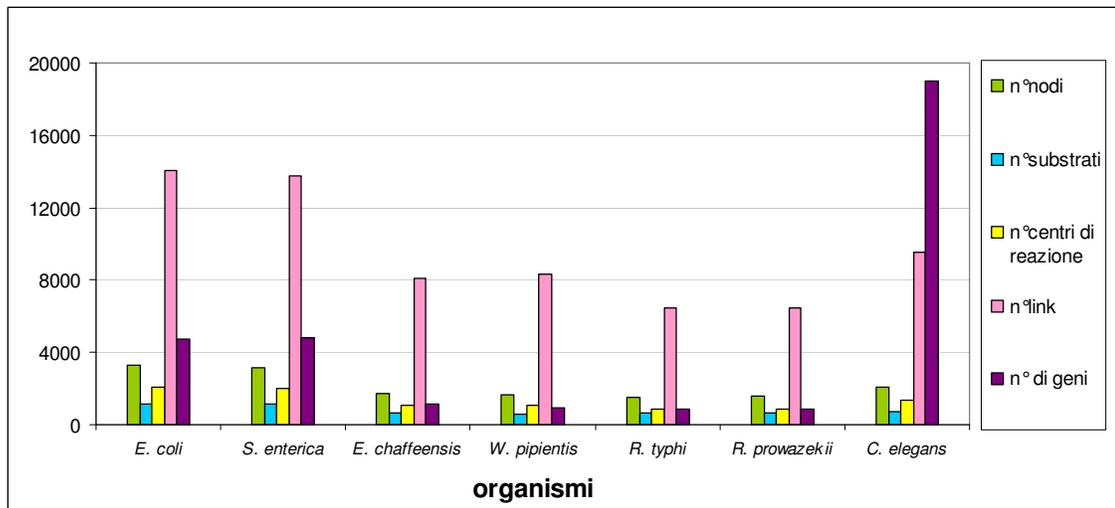


Fig. 17. Caratteristiche topologiche delle reti metaboliche degli organismi analizzati: numero di nodi, numero di substrati, numero di centri di reazione, numero di link e numero di geni.

Per quanto riguarda il numero di nodi si notano delle differenze legate allo stile di vita dei batteri: *E. coli* e *S. enterica*, batteri a vita libera, hanno una rete di dimensioni significativamente maggiori rispetto a quelle delle Rickettsiales, batteri intracellulari.

Le reti dei batteri intracellulari infatti hanno un numero di nodi che è circa la metà di quello dei batteri a vita libera; questa differenza è riconducibile alle dimensioni dei genomi e al numero di geni: un genoma più piccolo genera una rete più piccola.

Un genoma ridotto è una caratteristica tipica dei batteri intracellulari ed è il risultato del processo di *genome shrinkage* conseguente all'instaurarsi della simbiosi in cui vengono mantenuti solo quei geni necessari per uno stile di vita intracellulare: il processo '*Muller's ratchet*' (inevitabile fissazione di mutazioni deleterie in popolazioni a sessuate) è stato proposto come il fattore principale che porta alla degradazione del genoma negli organismi simbiotici obbligati (Pettersson e Berg, 2007). Come conseguenza della riduzione del genoma la rete metabolica risulta essere costituita da meno enzimi e meno substrati.

La rete di *C. elegans* ha dimensioni comparabili con quelle dei batteri; ciò è spiegabile con il fatto che essendo un organismo eucariote pluricellulare ha un metabolismo più complesso e quindi i dati presenti nel database si riferiscono al metabolismo basale comune a tutte le cellule del nematode; inoltre essendo il suo genoma di dimensioni elevate alcuni geni non sono ancora stati annotati.

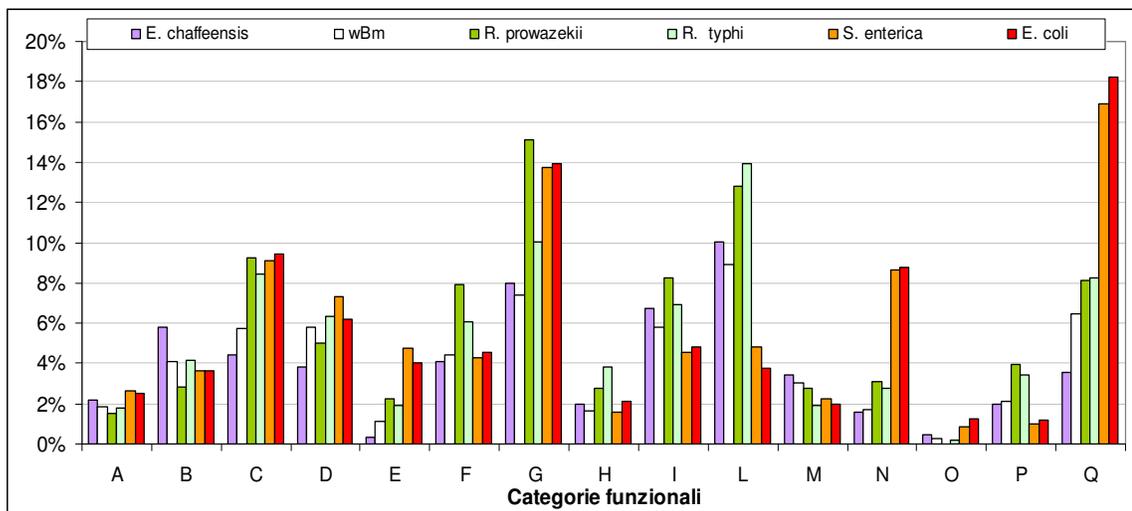
Come già detto i batteri a vita libera hanno un genoma di dimensioni molto maggiori rispetto agli intracellulari, in particolare si può osservare che i genomi di *E. coli* e di *S. enterica* sono circa quattro volte più grandi rispetto a quelli delle Rickettsiales considerate.

Per provare a mettere in relazione la dimensione della rete con le informazioni ottenute dalle analisi dei genomi è stato calcolato, per ciascuna specie, il rapporto tra il numero dei centri di reazione (CR) e il numero di geni (Tab. 5).

Tab. 5. Rapporto tra numero dei centri di reazione (CR) e numero di geni.

Organismo	Genoma (kb)	Geni	CR	CR/geni
<i>E. coli</i>	4938	4747	2088	0,44
<i>S. enterica</i>	4755	4801	2031	0,42
<i>E. chaffeensis</i>	1176	1115	1051	0,94
<i>R. typhi</i>	1111	838	888	1,06
<i>R. prowazekii</i>	1111	834	875	1,05
<i>W. pipientis</i>	1080	941	1100	1,17

Dai valori ottenuti si può osservare che negli organismi a vita libera il numero dei centri di reazione (CR) e il numero dei geni sono in rapporto 1:2, mentre negli intracellulari il rapporto è pari circa a 1:1; in altri termini i batteri a vita libera hanno un numero di geni che è doppio rispetto ai centri di reazione. A un numero di geni più piccolo, in rapporto, corrisponde una rete metabolica più grande. Sebbene gli organismi intracellulari vivendo all'interno della cellula-ospite abbiano a disposizione i metaboliti e quindi in generale abbiano perso parte dei pathway metabolici ciascuna specie è un caso a sé: le Anaplasmataceae hanno una percentuale significativamente elevata del genoma coinvolta nella biosintesi di nucleotidi, cofattori e vitamine e sintesi di proteine (Dunning Hotopp *et al.*, 2006); mentre le *Rickettsia* spp. nella sintesi delle proteine, metabolismo energetico e replicazione (Fuxelius *et al.*, 2007) (Fig. 18).



CATEGORIE FUNZIONALI			
<b>A</b>	Biosintesi degli amminoacidi	<b>I</b>	Protein fate
<b>B</b>	Biosintesi di cofattori e gruppi prostetici	<b>L</b>	Sintesi delle proteine
<b>C</b>	Sviluppo cellulare	<b>M</b>	Purine, pirimidine, nucleosidi e nucleotidi
<b>D</b>	Processi cellulari	<b>N</b>	Funzioni regolatorie
<b>E</b>	Metabolismo centrale intermedio	<b>O</b>	Trasduzione del segnale
<b>F</b>	Metabolismo DNA	<b>P</b>	Trascrizione
<b>G</b>	Metabolismo energetico	<b>Q</b>	Trasporto
<b>H</b>	Acidi grassi		

Fig. 18. Comparazione del contenuto genico (espresso come percentuale rispetto al genoma intero) degli organismi analizzati. I geni sono suddivisi nelle loro categorie funzionali, secondo la *TIGR annotation* (non sono state riportate le categorie particolarmente informative).

### **Diametro e distanza media**

Nei grafici in figura 19 sono riportati i valori misurati dei diametri e delle distanze medie in funzione della dimensione della rete; per entrambi i parametri non si osservano variazioni significative come è stato rilevato anche da Jeong *et al.* (2000); i dati invece non concordano con i valori dei diametri misurati da Ma e Zeng (2003) che osservano che il diametro aumenta con la dimensione della rete. Probabilmente quest'ultima discordanza di risultati è dovuta al fatto che sono state analizzate reti di organismi appartenenti a domini diversi quindi con una variabilità maggiore; invece in questo lavoro sono stati presi in esame sei specie di batteri filogeneticamente affini e una specie eucariote. Anche in questo caso i valori del diametro e della distanza media misurati in *C. elegans* sono molto simili a quelli misurati nei batteri: ciò è riconducibile alla struttura delle reti e alle dimensioni comparabili.

Con valori di distanza media compresi tra 3,97 e 4,6 passaggi le reti posso essere definite 'piccolo-mondo' in quanto nonostante l'elevato numero di nodi, qualsiasi coppia di substrati può essere connessa con un numero ridotto di passaggi.

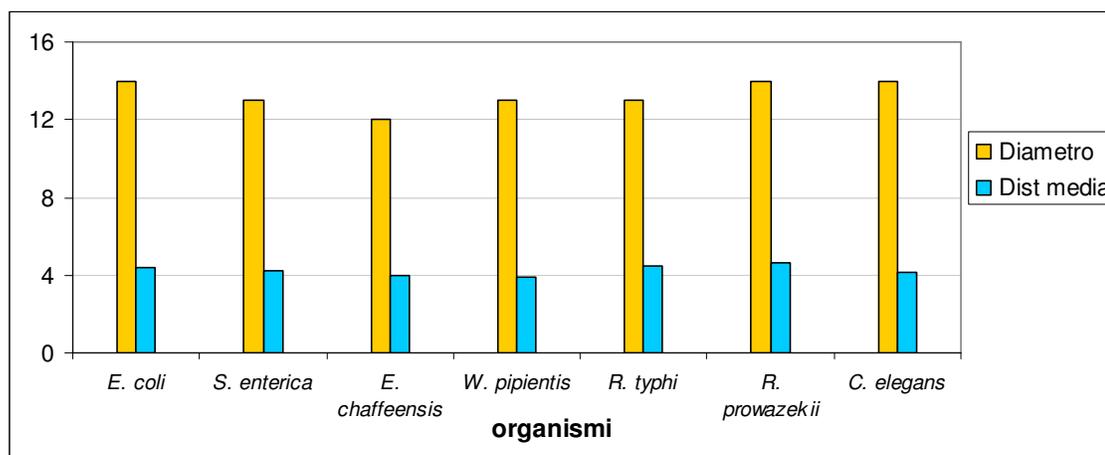


Fig. 19. Diametro e distanza media misurati per ciascun organismo.

### Distribuzione dei gradi dei nodi

I link delle reti ricostruite sono direzionati cioè si distinguono link in entrata ( $k_{in}$ ) e link in uscita ( $k_{out}$ ). Nei grafici in figura 20 sono mostrate separatamente le distribuzioni dei gradi dei nodi  $k_{in}$  e  $k_{out}$ , in scala logaritmica, della rete di *E. coli*, rappresentativa di tutte le reti dato che non sono state riscontrate differenze significative (Appendice A). Sull'asse delle ascisse sono riportati i gradi ( $k$ ), cioè il numero di link, mentre sull'asse delle ordinate è riportata la frequenza (o probabilità) di ciascun grado.

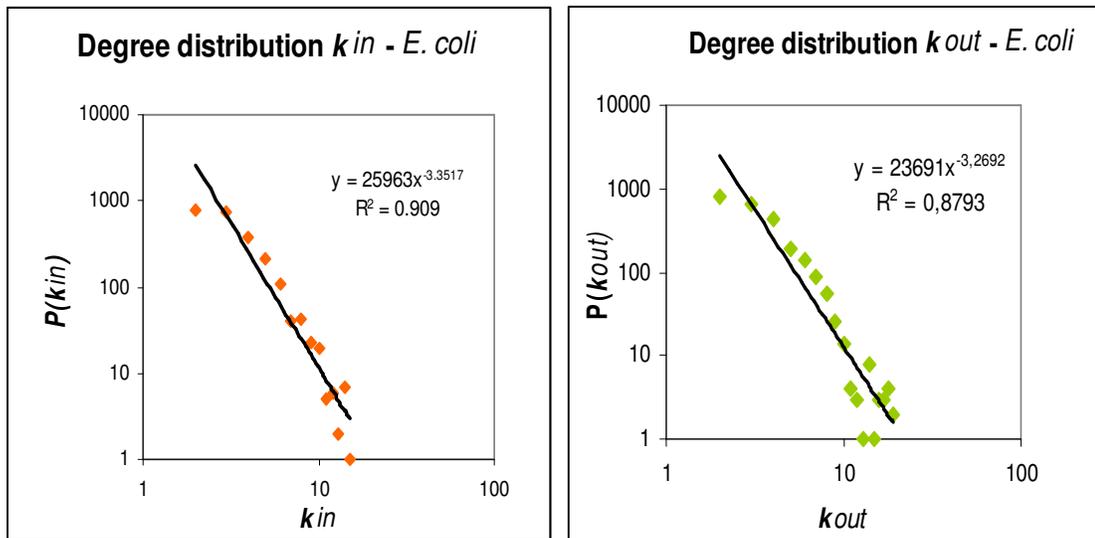


Fig. 20. - Distribuzione dei gradi dei nodi  $k_{in}$  e  $k_{out}$  della rete metabolica di *E. coli*.

Come si può osservare dal grafico, escludendo il punto  $k = 1$ , la probabilità  $P(k)$  di trovare un nodo con grado  $k$ , cioè con  $k$  link, è proporzionale a  $k^{-\gamma}$ . Ciò significa che la probabilità che un dato substrato partecipi come reagente in  $k$  reazioni segue una legge di potenza con valori di  $\gamma_{out}$  compresi tra 2,64 e 3,50, così come la probabilità che un dato substrato venga prodotto da  $k$  reazioni differenti segue una distribuzione simile con  $\gamma_{in}$  compresi tra 2,75 e 3,91.

L'esponente di grado delle due distribuzioni indica che ogni volta che il numero di link di un nodo raddoppia, il numero di nodi con quel grado diminuisce di più di 4 e 8 volte (cioè  $2^{2.64}$  e  $2^{3.50}$ ). Le reti metaboliche di tutti gli organismi quindi seguono il modello proposto da Barabási e Albert (1999) e si possono perciò definire reti a invarianza di scala.

### Identificazione degli hub

In una rete a invarianza di scala la coesistenza di hub e nodi con pochi link genera una legge di potenza nella distribuzione dei gradi di nodi, come descritto in precedenza. Gli hub sono i rari nodi altamente connessi che tengono insieme la rete; infatti è stato dimostrato che eliminando gli hub la rete si disgrega comportando un aumento della distanza tra i nodi e del

diametro; eliminando invece i nodi casualmente non si riscontrano effetti sulla connettività della rete (Albert, Jeong e Barabási, 2000). In ciascuna rete sono perciò stati identificati le molecole-hub che hanno un importante ruolo nel mantenere la struttura della rete metabolica.

Gli hub sono stati identificati in base alla definizione proposta da Wagner e Fell (2001): un nodo con un numero di link ( $k$ ) maggiore della somma tra la media dei gradi dei nodi ( $k_m$ ) e tre volte la deviazione standard è considerato hub, cioè  $k > k_m + 3\sigma$ .

I nodi identificati come hub in tutte le reti sono **ATP, H<sub>2</sub>O, ADP, ortofosfato, NADH, CO<sub>2</sub>, piruvato, NADPH, NADP<sup>+</sup>, CoA**. Come è noto, sono tutte molecole con un ruolo chiave nel metabolismo dato che partecipano a quasi tutte le reazioni metaboliche, per esempio il piruvato costituisce uno dei composti intermedi della glicolisi; NAD, NADP<sup>+</sup> e NADPH sono importanti coenzimi coinvolti in molteplici reazioni di ossidoriduzione; il CoA è un coenzima tiolico trasportatore di gruppi acili od acetili, interviene in numerosi processi biochimici: entra nelle vie metaboliche dei carboidrati, degli amminoacidi, degli acidi grassi, dei composti steroidei, dei corpi chetonici e partecipa anche ad altre reazioni di acetilazione.

Queste stesse molecole sono state identificate come hub anche in lavori precedenti condotti sulle reti metaboliche di numerosi altri organismi nei quali sono stati infatti definiti ‘*key metabolites*’ (Wagner e Fell, 2001) o ‘*current metabolites*’ (Ma e Zeng, 2003).

Non sono state riscontrate grosse differenze tra le molecole hub di *C. elegans* rispetto a quelle dei batteri.

Osservando l’elenco degli hub non presenti in tutti gli organismi (Tab. 6), invece, si riscontrano maggiori differenze.

Tab. 6. Elenco degli hub non presenti in tutti gli organismi; il simbolo ‘+’ indica che la molecola è stata identificata come hub nella rete dell’organismo, il simbolo ‘-’ indica che la molecola non è un hub.

HUB	<i>E. co</i>	<i>S. en</i>	<i>E. ch</i>	wBm	<i>R. ty</i>	<i>R. pr</i>	<i>C. el</i>
NAD <sup>+</sup>	+	+	+	+	+	+	-
AMP	+	+	+	-	+	+	-
D-glucosio	+	+	+	+	-	-	+
L-glutammato	+	+	-	-	+	+	+
NH <sub>3</sub>	+	+	-	-	-	-	-
2-oxoglutarato	+	+	-	-	+	+	+
Acetil-CoA	+	+	-	-	+	+	-
H <sup>+</sup>	-	+	-	-	-	-	-
O <sub>2</sub>	-	-	-	+	-	-	-
Pirofosfato	+	+	+	-	+	+	+

*E. co*, *Escherichia coli*; *S. en*, *Salmonella enterica*; *E. ch*, *Ehrlichia chaffeensis*; wBm, *W. pipientis* endosimbionte di *B. malayi*; *R. ty*, *Rickettsia typhi*; *R. pr*, *R. prowazekii*; *C. el*, *Caenorhabditis elegans*.

I due batteri a vita libera, *E. coli* e *S. enterica*, presentano gli stessi hub tranne che per lo ione  $H^+$ . Gli altri organismi invece hanno un numero minore di hub; questo è legato ai parametri calcolati per definire gli hub e quindi alle dimensioni delle reti e al grado dei nodi.

Le molecole presenti nell'elenco hanno tutte un ruolo fondamentale nel metabolismo infatti acetil-CoA è il metabolita che connette il pathway glicolitico, il ciclo dell'acido citrico e quello per la sintesi degli acidi grassi; L-glutammato è un importante amminoacido prodotto direttamente dal precursore e che può essere convertito in altri amminoacidi; il 2-oxoglutarato trasferisce gruppi amminici nelle reazioni.

## **2. Analisi dell'interazione ospite-simbionte**

Al fine di studiare l'interazione tra i microrganismi endosimbionti e i loro ospiti sono state individuate quelle molecole che risultano tra i 'solo prodotti' e tra i 'solo reagenti' delle reazioni.

Scopo del lavoro è analizzare l'interazione ospite-simbionte nel caso particolare di *Wolbachia pipientis* endosimbionte del nematode parassita *Brugia malayi* e confrontare le molecole coinvolte nell'interazione tra organismi patogeni e ospite.

Il tipo di relazione che coinvolge *W. pipientis* e *B. malayi* è di tipo mutualistico, si vuole quindi studiare questa interazione individuando quelle molecole oggetto di scambio tra i due organismi. Dato che il genoma di *B. malayi* è stato sequenziato solo di recente (Ghedini *et al.*, 2007) e quindi il suo metabolismo non è ancora disponibile nel database PUMA2, si è deciso di utilizzare come organismo modello per il confronto un altro nematode ben studiato e cioè *C. elegans*. Sono state quindi individuate le potenziali molecole di interazione; sono stati messi a confronto i metaboliti che risultano tra i 'solo prodotti' di *W. pipientis* con le molecole che compaiono tra i 'solo reagenti' dell'ospite nematode e, allo stesso modo, sono state confrontate le liste dei 'solo reagenti' del batterio con i 'solo prodotti' del nematode.

Infine sono state analizzate le molecole di interazione che vengono acquisite e secrete dai microrganismi patogeni; nel confronto è stata inclusa anche *W. pipientis* che, sebbene svolga un ruolo da mutualista nella relazione con il nematode, è stato dimostrato che nell'uomo contribuisce alla patogenesi della filariosi. Sono state quindi confrontate le molecole di interazione di *W. pipientis* con quelle delle altre specie di microrganismi patogeni al fine di individuare somiglianze e differenze.

## 2. 1. Interazione *Wolbachia pipientis* – nematode

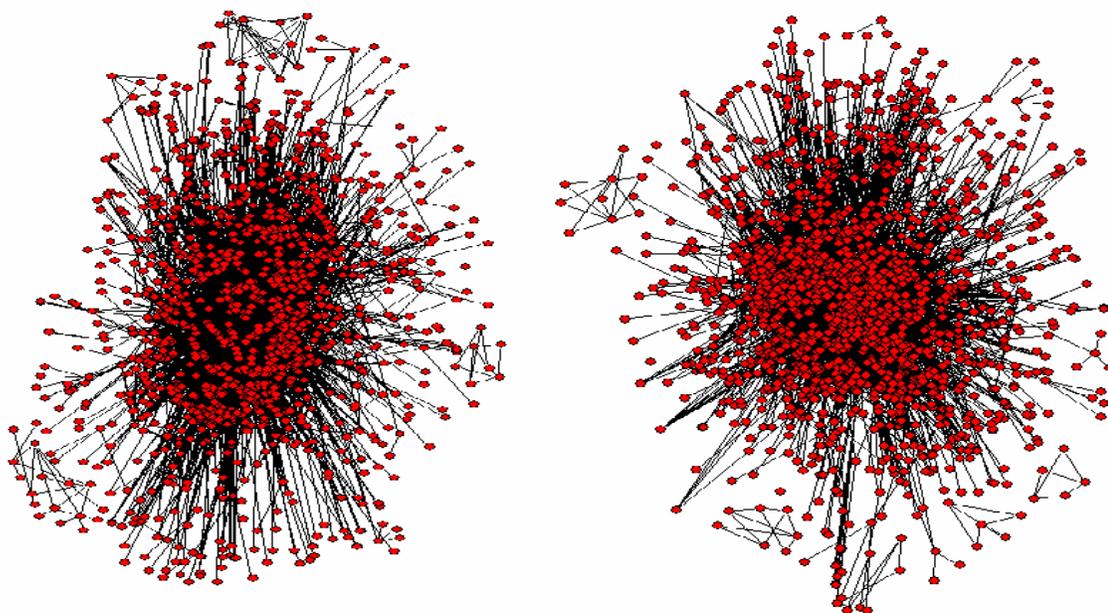


Fig. 21. Rete metabolica di *W. pipientis* (wBm) (a sinistra) e rete metabolica di *Caenorhabditis elegans* (a destra). Le rete sono state disegnate utilizzando il programma Pajek (Batagelj e Mrvar, 1998).

Tab. 7. Caratteristiche topologiche della rete metabolica di *W. pipientis* (wBm) e di *Caenorhabditis elegans*

Caratteristiche topologiche	<i>W. pipientis</i>	<i>C. elegans</i>
<b>Numero di nodi</b>	1776	2101
<b>Numero di substrati</b>	676	750
<b>Numero di centri di reazione</b>	1100	1351
<b>Diametro</b>	13	14
<b>Densità</b> (numero di nodi/diametro)	136,62	150,07
<b>Distanza media tra i nodi</b> (gradi di separazione)	4,03	4,15
$\gamma_{in}$	2,99	3,14
$\gamma_{out}$	2,73	3,00

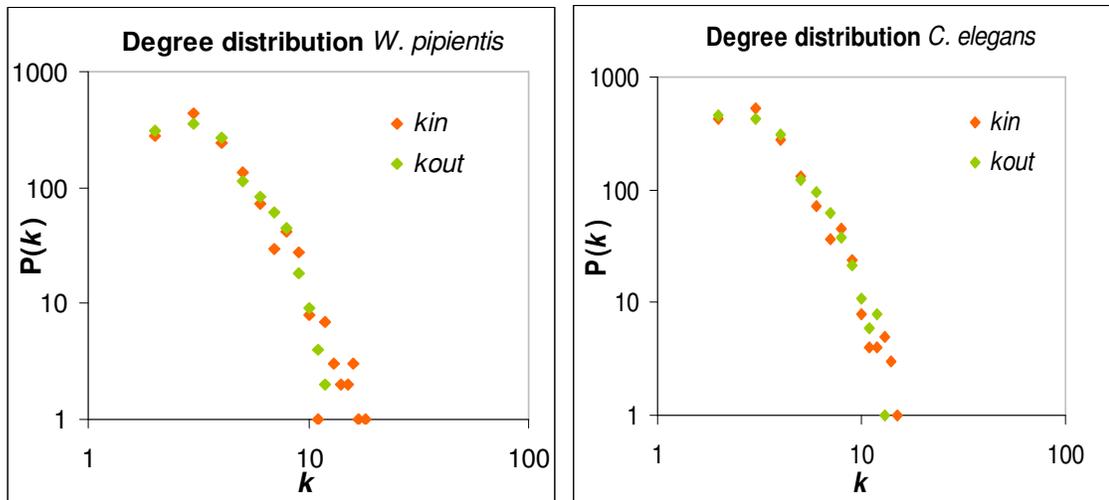


Fig. 22. Distribuzione dei gradi dei nodi delle reti metaboliche di *W. pipientis* e di *C. elegans*.

Nella figura 22 sono mostrate separatamente le distribuzioni dei gradi dei nodi di *W. pipientis* e *C. elegans*. Entrambe le reti metaboliche possono essere definite reti piccolo-mondo, hanno cioè pochi gradi di separazione (Tab. 7) e seguono il modello a invarianza di scala.

*W. pipientis* e i nematodi filaridi hanno instaurato un mutualismo obbligato in cui si scambiano composti necessari per la vita di entrambi: dall'analisi del genoma di *W. pipientis* è emerso che le sue capacità biosintetiche sono piuttosto limitate per questo motivo deve ottenere dall'ospite alcune molecole essenziali per la sopravvivenza; così come l'analisi del genoma di *B. malayi* ha mostrato che i set di geni per la sintesi di alcune molecole essenziali sono del tutto incompleti.

I risultati del confronto sono riportati nelle tabelle 8 e 9, dove sono state messe in evidenza le molecole in comune.

Tab. 8. Confronto tra le molecole che escono da *W. pipientis* e le molecole che entrano in *C. elegans*; in blu sono evidenziate le molecole in comune.

	<b>‘Solo prodotti’ di <i>W. pipientis</i> (wBm)</b>	<b>‘Solo reagenti’ di <i>C. elegans</i></b>
<b><u>Vitamine e cofattori</u></b>	Mioinositolo ( <i>vit B</i> ), <b>diidrofolato</b> ( <i>vit B9</i> ), <b>piridoxammina</b> ( <i>vit B6</i> ), adenosilcobalammina ( <i>vit B12</i> ), diidrometanopterinina, tetraidrometanopterinina, FADH, FADH <sub>2</sub> , menaquinone ( <i>K</i> ), molibdopterina, protoeme, siroeme, <b>coenzima F420 ridotto</b> , <b>pirrolochinoline-chinone ridotto</b> (PQQred), ubiquinone (coQ), palmitate ( <i>vit A</i> )	FAD <sup>+</sup> , ferrodossina, ferrocitocromo c, <b>coenzima F420</b> , acil-CoA, lipoamide, <b>piridoxammina</b> ( <i>vit B6</i> ), <b>pirrolochinoline-chinone</b> (PQQ), nicotina, <b>10-formiltetraidrofolato</b> , <b>5-metiltetraidrofolato</b> , <b>7,8-diidrofolato</b> , uroporphyrinogeno III
<b><u>Nucleotidi</u></b>	AMP, CMP, GMP, UMP, dITP, <b>ITP</b> , NTP, GDP-6-deossi-D-mannosio, GDP-D-ramnosio, GDP-fucosio, GDPmannosio, adenina, adenosina, guanina, <b>guanosina</b> , ipoxantina, inosina, pUpU, nicotinamide nucleotide, 25,10-metilenetetraidrofolato, TTP ( <i>timidina trifosfato</i> )	CTP, <b>ITP</b> , TTP, UTP, XTP, cAMP, cGMP, dAMP, dCMP, <b>guanosina</b> , uridina, piridina, timina, citidine, deossiadenosina, deossitimidina
<b><u>Glicolisi/ TCA</u></b>	Lattato, malato, alfaDglucosio1,6biP, <b>alfaDribosio1P</b> , glutarato, butanolo, catecolo, etanolo, succinato, S-lattaldeide, S-piruvato, acetaldeide, <b>alfaDglucosio</b> , fruttosio6P	Isocitrato, metanolo, alfaDglucosio6P, alfaDglucosio1P, <b>alfaDribose1P</b> , <b>alfaDglucosio</b> , betaDglucosio, D-mioinositolo 1,4,5-trisfosfato, D-gluconato, isocitrato, fosfoenolpiruvato
<b><u>Metabolismo dei carboidrati</u></b>	Saccarosio	Xilitolo, sorbitolo, ribosio, L-xilose, L-arabitol, L-ramnosio, lattosio, glycogen amylose chain, D-mannosio, D-sorbosio, D-galattosio, D-fruttosio, D-arabinose, D-arabinitol, L-lyxosio
<b><u>Metabolismo degli acidi grassi</u></b>	(N <sup>-</sup> -acetil-D-glucosaminil-N <sup>-</sup> -acetilmuramoil-pentapeptidil)-diphosphoundecaprenol, acetone, butanoil-CoA, S-3-idrossi-3-metilglutaril-CoA, S-acetil-CoA, lipoproteina	oleoil-CoA, colesterolo, colestene, tetradecanoil-CoA
<b><u>aa</u></b>	Glutammato, ornitina, <b>prolina</b> , tiocisteina, diidrolipoilproteina, L-omocisteina	<b>D-prolina</b> , D-glutammina, D-glucosammina, L-allotreonina, L-asparagina, L-aspartato-semialdeide, L-guanidinobutirate, L-omoserina, L-triptofano, L-aspartato semialdeide, L-guanidinobutirato, D-alanina, corismato, D-aspartato, omoserina, sarcosina
<b><u>altro</u></b>	Clorofilla a, CO <sub>4</sub> , <b>CO<sub>2</sub></b> , H <sub>2</sub> O, urea, carbamato, glutatione, initiation factor IF-2-GDP, reduced flavodossina, glutaredossina ossidata, tioredossina ossidata, glutatione	<b>CO<sub>2</sub></b> , nicotina, Mg <sup>2+</sup> , selenide

Tab. 9. Confronto tra le molecole che entrano in *W. pipientis* e le molecole che escono da *C. elegans*; in blu sono evidenziate le molecole in comune.

	<b>‘Solo reagenti’ di <i>W. pipientis</i> (<i>wBm</i>)</b>	<b>‘Solo prodotti’ di <i>C. elegans</i></b>
<b><u>Vitamine e cofattori</u></b>	Ferrodissina, coenzima F420, FAD, coproporfirinogeno III, pirrolochinoline-chinone ( <i>PQQ</i> ), quinate, TMP ( <i>vit B1</i> ), piridoxal fosfato, <b>5-metiltetraidrofolato</b> , uroporfirinogeno III	Enterobactin, FADH, FADH <sub>2</sub> , ferrocitocromo c, mioinositolo, manachinone, ubichinone, diidrofolato ( <i>vit B9</i> ), <b>5-metiltetraidrofolato</b> , nicotinate ( <i>vit B3</i> )
<b><u>Nucleotidi</u></b>	IDP, dAMP, dGMP, <b>dGDP</b> , dIDP, <b>dCDP</b> , dCMP, deossadenosina, deossimidina, deossicitidina, , queuine	AMP, CMP, GMP, UMP, TDP, XDP, NADP, dADP, dCDP, <b>dTDP</b> , dUDP, <b>dGDP</b> , pUpU, adenosina
<b><u>Glicolisi/ TCA</u></b>	<b>AlfaDglucosio</b> , <b>betaDglucosio</b> , D-gluconato, isocitrato, malonato, formaldeide	<b>AlfaDglucosio</b> , alfaDglucosio 1,6diP, Dglucosio, acetoacetato, D-glucosio 1,6biP, butanolo, carbamato, catecolo, D-pinitolo, fumarato, glicolaldeide, glicolato, lattato, acetyl-Coa, <b>betaDglucosio</b> , formato, L-ascorbato, L-xylonate, propanoate, succinato
<b><u>Metabolismo dei carboidrati</u></b>	D-fruttosio, D-galattosio, D-mannosio, L-ramnosio, ribosio, D-arabinitolo, ribitolo, sorbitolo, xylitolo, alfa-D-glucose 1P, UDPglucosio	D-xylulosio 5P, saccarosio
<b><u>Metabolismo degli acidi grassi</u></b>	Glicerone, acil-CoA, linoleoil-CoA, oleoil-CoA, palmitoil-CoA, arachidonoil-CoA, oleoil-CoA	Butanoil-CoA
<b><u>aa</u></b>	L-asparagine, Beta-alanina, corismato, D-alanina, <b>D-glutamato</b> , <b>D-prolina</b> , L-allotreonina, L-asparagina, L-treonina, L-triptofano, carbamoil fosfato, L-aspartato semialdeide, L-aspartato 4-semialdeide, lipoilproteina	<b>Glutammato</b> , <b>prolina</b>
<b><u>Altro</u></b>	H, O <sub>2</sub> , Mg <sup>2+</sup> , HCO <sub>3</sub> <sup>-</sup> , host O-sialoglicoproteina, flavodossina ossidata, glutatione ossidato, porfirinogeno IX, glutaredossina ridotta, tioredossina ridotta	CO <sub>2</sub> , CO <sub>4</sub> , Fe <sup>2+</sup> , urea, zeaxantina diglucoside, gramicidin S, serotonina, spermina, cadaverina, sferoidene, taurina, ubiquitina

*W. pipientis* sembra acquisire buona parte degli **amminoacidi** dall'esterno: D-alanina, L-allotreonina, L-asparagina, L-treonina, D-prolina, L-triptofano. Inoltre importa il corismato, un importante precursore degli amminoacidi; il D-glutammato che viene sintetizzato direttamente dal precursore e può essere convertito in molti altri amminoacidi e L-aspartato 4-semialdeide, un intermedio nella biosintesi dell'omoserina e lisina. I dati genomici confermano i risultati ottenuti: *wBm* è in grado di sintetizzare un solo amminoacido, il *meso*-DAP, perciò gli amminoacidi devono essere importati dall'esterno e sono infatti ottenuti in seguito alla proteolisi delle proteine dell'ospite (Foster *et al.*, 2005).

I nematodi, come i mammiferi, devono ottenere molti degli amminoacidi essenziali dalla dieta quindi *C. elegans*, come *W. pipientis*, deve importarli dall'esterno e quindi dall'ospite, infatti

molti amminoacidi risultano in entrata. Gli amminoacidi sono poi utilizzati dal nematode per sintetizzare le proteine che a loro volta possono essere digerite da enzimi batterici e assunti da *W. pipientis*.

*W. pipientis* sembra acquisire i **nucleotidi** monofosfato, dGMP, dAMP, dCMP, i nucleosidi deossiadenosina, deossicitidina, deossitimidina e pochi altri; invece i nucleotidi in uscita dal batterio sono in numero maggiore, a suggerire che questi vengono sintetizzati per il consumo interno ma anche forniti all'ospite, infatti alcuni nucleotidi risultano essere assunti da *C. elegans*. Ciò coincide anche a quanto emerso dall'analisi del genoma di *wBm*: sono presenti i pathway completi per la biosintesi *de novo* di purine e pirimidine e sembra che vengano assunti dall'ospite nematode nei momenti in cui è maggiore la richiesta di nucleotidi per la sintesi di DNA e cioè durante l'oogenesi e l'embriogenesi (Foster *et al.*, 2005). Nel caso particolare, in *B. malayi* i set di geni per la sintesi di purine sono quasi del tutto incompleti, rendendolo perciò dipendente da *Wolbachia* per questi nucleotidi.

Tra i **cofattori e vitamine** che *W. pipientis* acquisisce dall'esterno ci sono coenzima F420, pirrolochinoline-chinone, quinate e TMP (vitamina B<sub>1</sub>), ferrodossina, pantotenato, pantoato, FAD, piridoxal fosfato e folato. In *wBm* sono effettivamente assenti i pathway per tiamina, folato, piridoxina fosfato; oltre a questi appena citati tra i composti che *wBm* non è in grado di sintetizzare *de novo* ci sono anche biotina, ubiquinone, acido lipoico e NAD; di questi composti in realtà sembra che *wBm* abbia mantenuto alcuni degli step finali quindi è possibile che il batterio ottenga dall'ospite i precursori, per questo motivo potrebbero non comparire tra le molecole in entrata. Per esempio è completamente assente il pathway per la sintesi del coenzima A che però non risulta in entrata, ma *wBm* acquisisce il pantotenato che insieme all'ATP e beta-mercaptoetilamina concorre alla formazione del coenzima A. Inoltre diverse forme o precursori di folato, ubichinone e di altri cofattori vengono prodotti dal nematode e secreti verso l'esterno e risultano anche in entrata in *W. pipientis* dove potrebbero essere presenti gli step finali del percorso di biosintesi e quindi gli permette di ottenere la forma attiva del coenzima.

Il nematode a sua volta dipende da *W. pipientis* per altri composti che non è in grado di sintetizzare; è il caso per esempio di due importanti intermedi necessari per la sintesi dell'eme: protome e siroeme risultano in uscita da *W. pipientis* e a sua volta *C. elegans* acquisisce dall'esterno l'uroporfirinogeno III, un intermedio prodotto durante la sintesi dell'eme. L'eme oltre a prendere parte a numerosi processi biologici come la detossificazione, il metabolismo ossidativo e la differenziazione cellulare (Rao *et al.*, 2005) sembra giocare un ruolo fondamentale durante la riproduzione e lo sviluppo della filaria (Foster *et al.*, 2005).

Diversi studi condotti a livello genomico, di biochimica e sulle capacità nutrizionali hanno dimostrato che *C. elegans* non è in grado di sintetizzare l'eme (Rao *et al.*, 2005) e in *B. malayi* è completamente perso il pathway di sintesi, mentre in *W. pipientis* è presente il set di geni completo.

La vitamina B<sub>12</sub> (adenosil-cobalamina), è un'importante cofattore che in *W. pipientis* risulta essere secreto verso l'esterno quindi probabilmente è in grado di sintetizzarla e allo stesso tempo di fornirla all'ospite: solo i batteri infatti sono in grado di produrla e quindi, come tutti gli altri animali, anche i nematodi devono assumerla dall'esterno.

Tra i prodotti di *W. pipientis* compare il palmitate (o vitamina A o retinolo); dati bibliografici riportano che *B. malayi* assorbe retinolo o acido retinoico direttamente dal mezzo di coltura oltre a essere importante nello sviluppo, riproduzione e metabolismo delle filarie parassite. *W. pipientis* quindi, oltre all'ospite vertebrato, potrebbe rappresentare per *B. malayi* una fonte 'sicura' di nucleotidi.

FADH e FADH<sub>2</sub> vengono prodotti da *W. pipientis* mentre *C. elegans* acquisisce dall'esterno e utilizza quindi come reagente il FAD<sup>+</sup>; in *wBm* è presente il pathway per la sintesi di riboflavina (vitamina B<sub>2</sub>) e FAD (cioè la forma attiva della vitamina B<sub>2</sub>) mentre in *B. malayi* è quasi totalmente incompleto il set di geni per la biosintesi della riboflavina, perciò potrebbe acquisire dal batterio uno degli intermedi finali necessari per la sintesi di riboflavina e FAD.

Il glutatione compare tra i prodotti in uscita da *W. pipientis*, è un metabolita essenziale contro lo stress ossidativo e per la detossificazione. Sebbene siano presenti i geni per la sua biosintesi sia in *W. pipientis* che in *B. malayi* è possibile che in condizioni di stress *wBm* fornisca il glutatione all'ospite elevato.

*W. pipientis* vivendo all'interno del corpo dell'ospite utilizza come substrati di crescita quei composti che sono abbondanti nel nematode. Diverse molecole che costituiscono gli intermedi del TCA risultano infatti in entrata in *W. pipientis* come ad esempio isocitrato, malonato, glucosio. Succinato, lattato sono i principali prodotti finali del processo di utilizzazione del glucosio che insieme ad altri intermedi del ciclo dei TCA, come piruvato, acetilCoA, fumarato e fruttosio, vengono secreti dal nematode e potrebbero dunque essere utilizzati da *W. pipientis* come composti per la crescita. Inoltre sembra che *C. elegans* sia in grado di produrre due molecole che rientrano nella composizione del peptidoglicano (N-acetilneuraminate e N-acetilglucosammina) che dunque potrebbero essere acquisite dal batterio per sintetizzare la parete.

In generale tra le varie molecole di natura polisaccaridica secrete da *W. pipientis* risultano lipopolisaccaridi, fosfolipidi, lipidi che costituiscono i principali componenti della membrana

e della parete. Da bibliografia è noto che sia *C. elegans* che *B. malayi* mancano delle proteine implicate nel riconoscimento del peptidoglicano e dell'LPS (Ghedin *et al.*, 2007), questo potrebbe spiegare perché in *B. malayi* la presenza di *W. pipientis* non induca una reazione negativa mentre nel mammifero il rilascio del batterio stesso o dei suoi prodotti scatena la risposta infiammatoria.

Nel confronto condotto tra *W. pipientis* e *C. elegans* non è stata riscontrata una corrispondenza perfetta tra i composti in entrata e quelli in uscita (e viceversa) dei due organismi. Tale incongruenza è spiegabile innanzitutto da errori presenti nei dati di partenza e cioè nel database da cui sono stati scaricati ma anche errori commessi durante la correzione delle reti metaboliche (entrambi sono difficile da evitare).

In secondo luogo per lo studio della simbiosi in cui è coinvolta *W. pipientis*, per motivi di disponibilità di dati, è stato condotto utilizzando la rete metabolica di *C. elegans*. Infatti come detto in precedenza la rete metabolica di *B. malayi*, l'ospite mutualista di *W. pipientis*, non è ancora disponibile nel database. Sebbene sia stato scelto un altro nematode molto ben studiato per il confronto, bisogna sottolineare il fatto che *C. elegans* è un nematode a vita libera mentre *B. malayi* è un nematode filaride parassita; inoltre tra *B. malayi* e *W. pipientis* si è instaurata una relazione molto intima, mutualismo obbligato, che li rende dipendenti uno dall'altro. *C. elegans* invece ha mantenuto la maggior parte dei pathway per la sintesi di composti essenziali quelli cioè che non riesce ottenere dall'ambiente esterno, ciò è legato essenzialmente allo stile di vita condotto. Quindi dati i due diversi stili di vita, le loro esigenze metaboliche possono risultare estremamente differenti.

Diversi lavori hanno dimostrato che è possibile un trasferimento di informazione tra organismi diversi mediato da RNA: Fire *et al.* (1998) hanno dimostrato che attraverso esperimenti di *RNA interference* (RNAi) è possibile inibire l'espressione di geni target in *C. elegans* tramite iniezione di dsRNA. Lo stesso effetto si è osservato nutrendo *C. elegans* con *E. coli* che trascrivono *doublestranded RNA* (dsRNA) ricombinante. Tale meccanismo potrebbe essere implicato nelle interazioni ecologiche e in particolare nei casi di simbiosi (Timmons e Fire, 1998).

## 2. 2. Molecole di interazione negli organismi patogeni

Nel parassitismo l'ospite è svantaggiato dalla presenza del parassita che manipola o sfrutta le sue risorse; inoltre in un'infezione batterica la prima reazione che si registra nell'organismo-ospite è l'originarsi di un processo infiammatorio: i composti responsabili dell'attivazione delle cellule del sistema immunitario dei vertebrati sono i principali componenti delle strutture esterne dei batteri: sono molecole di natura polisaccaridica come i lipopolisaccaridi (o LPS), o di natura polipeptidica.

I risultati dell'analisi sono riportati in appendice B e C.

Diversi composti che costituiscono gli intermedi del **metabolismo dei carboidrati/TCA** (quali per esempio isocitrato, malonato, glucosio) risultano essere assunti da tutte e sei le specie di batteri considerate ma differiscono per alcuni composti:

- piruvato, PEP, acetil-CoA, Dfruttosio 6P si trovano solo tra i reagenti di *R. prowazekii*, mentre il glicerolo solo in *R. typhi*: le specie del genere *Rickettsia* non possiedono la via glicolitica e devono perciò ottenere questi substrati dalla cellula ospite.
- glicerone fosfato, 2P-D-glicerato, e gliceraldeide 3P risultano in entrata in *E. chaffeensis* e *W. pipientis*.

*E. coli* e *S. enterica* oltre ad acquisire le stesse molecole che risultano in entrata nelle specie di Rickettsiales, assumono anche altri composti intermedi del TCA come succinil-CoA e citrato.

Queste molecole sono importanti substrati di crescita per i batteri per cui è ragionevole pensare che siano presenti nella cellula-ospite o nelle cavità dell'ospite e vengano perciò acquisiti dai parassiti non appena disponibili.

Confrontando la lista degli **amminoacidi** in entrata e in uscita dai batteri si può osservare una significativa differenza: è estremamente ridotto il numero degli amminoacidi che vengono secreti dai batteri mentre un numero significativamente elevato di amminoacidi risulta invece in entrata. Sono sia amminoacidi nella loro struttura definitiva ma anche precursori e composti intermedi; per esempio corismato (precursore), D-alanina, L-allotreonina, L-asparagina, L-treonina, prolina, D-glutammato, L-aspartato semialdeide e L-aspartato 4-semialdeide (intermedi nella biosintesi dell'omoserina e lisina) sono presenti nella lista dei 'solo reagenti' di tutte le Rickettsiales considerate, a suggerire che, probabilmente a partire da questi amminoacidi, acquisiti dall'ospite, i batteri sono in grado di ottenere la maggior parte degli altri amminoacidi necessari per la sopravvivenza. I risultati riguardanti l'elevato numero di amminoacidi in entrata nelle specie di Rickettsiales trovano riscontro nei dati ottenuti

dall'analisi del genoma; le Rickettsiales infatti hanno perso la maggior parte dei set di geni necessari per la sintesi degli amminoacidi devono perciò acquisirli dalla cellula-ospite.

Anche *E. coli* e *S. enterica* acquisiscono un elevato numero di amminoacidi, però differiscono dalle Rickettsiae per altri importanti amminoacidi. Non risultano infatti avere in entrata il corismato (precursore degli amminoacidi) e il D-glutammato, hanno invece isocorismato, glutammato e D-aspartato tra i composti in uscita quindi i pathway per la sintesi di questi amminoacidi sarebbero presenti e a partire da essi potrebbero, in caso di necessità, ottenere gli altri: aspartato e glutammato infatti sono due importanti amminoacidi direttamente prodotti dai precursori che possono essere convertiti in molti altri.

**Cofattori e vitamine** sono necessari per l'azione di certi enzimi e vengono sia secreti verso l'esterno che importati dalle cellule batteriche.

Tra i cofattori e vitamine che le sei specie di batteri acquisiscono dall'esterno ci sono: coenzima F420, coproporfirinogeno III, pirrolochinoline-chinone, quinate e TMP (vitamina B<sub>1</sub>), ferrodossina, pantotenato e pantoato.

Dai dati disponibili dall'analisi dei genomi è noto che i pathway per tiamina e pantotenate nelle Rickettsiales sono assenti. Si osserva una possibile incongruenza solo con *E. chaffeensis*, dato che sembra essere in grado di sintetizzare la maggior parte di vitamine e cofattori e quindi non dovrebbe dipendere per queste molecole dall'ospite. Nei casi specifici:

- *R. prowazekii* risulta acquisire quei coenzimi che non è in grado di sintetizzare: FAD, NAD, coenzima A, folato, piridoxina-fosfato (vitamina B<sub>6</sub>);

- *R. typhi* acquisirebbe il NAD.

Il NAD risulta in entrata, infatti il genere *Rickettsia* non è in grado di sintetizzarlo: mancano i geni per la sintesi di nicotinate e nicotinamine (i precursori di NAD<sup>+</sup> e NADP<sup>+</sup>) e sono però presenti nel genoma gli enzimi coinvolti nell'interconversione tra NAD<sup>+</sup> e NADP<sup>+</sup>. Acquisiscono inoltre la nicotina i cui anelli pirimidinici sono utilizzati per la sintesi di NAD e NADP.

- in *wBm* risultano, inoltre, in entrata FAD, piridoxina-fosfato (B<sub>6</sub>) e folato i cui pathway sono andati persi.

Protoeme, siroeme, menachinone, tetraidrometanopterin, molibdopterina, coenzima F420 ridotto, FADH, diidrofolato (vitamina B<sub>9</sub>), ubiquinone, diidrometanopterin, pirrolochinoline-chinone ridotto, palmitate sono tutti cofattori che vengono prodotti dalle sei specie di batteri.

In *W. pipientis* protoeme e siroeme effettivamente sono importanti composti, essenziali nella relazione mutualistica con l'ospite-nematode: l'eme infatti sembra svolgere un ruolo

fondamentale per la riproduzione e lo sviluppo di *B. malayi* la quale ha perso il pathway di sintesi.

La biotina si trova solo tra i prodotti di *E. chaffeensis* e dei due batteri a vita libera *E. coli* e *S. enterica*; è un cofattore essenziale coinvolto in reazioni di carbossilazione ed è comunemente sintetizzato dai batteri intestinali e dalle specie di Anaplasmataceae in grado di infettare vertebrati. Diversi ioni inorganici, alcuni dei quali svolgono il ruolo di cofattori di enzimi, risultano essere acquisiti e secreti dai batteri:  $Mg^{2+}$ ,  $HCO_3^-$ ,  $Co^{2+}$ ,  $Fe^{2+}$ ,  $Na^+$ .

Tutti i batteri analizzati importano prevalentemente **nucleotidi** monofosfato e deossinucleotidi oltre ad altri che però variano da specie a specie. Si può osservare una differenza significativa tra i batteri del genere *Rickettsia* e gli altri analizzati: *R. typhi* e *R. prowazekii* primi risultano assumere un maggior numero di nucleotidi diversi a conferma del fatto che in questi batteri è avvenuta una drastica riduzione dei pathway per la sintesi *de novo* dei nucleotidi; sono stati mantenuti però i geni per l'interconversione di nucleosidi monofosfato in tutti gli altri nucleotidi.

Per la sintesi dei **polisaccaridi extracellulari** i batteri analizzati sembrano importare dall'esterno le molecole base costituenti i polimeri: UDPglucosio, D-glutammato, D-alanina, L-alanina, L-lisina e N-acetil-muramico sono alcune delle molecole che rientrano nella composizione del peptidoglicano, il principale componente della parete cellulare.

Le due specie di Anaplasmataceae si distinguono rispetto alle altre specie: *E. chaffeensis* non ha in entrata N-acetil-muramico infatti nel suo genoma non sono presenti i geni per sintetizzare il peptidoglicano, mentre *W. pipientis* non importa L-lisina che però può essere sostituita dal DAP che è in grado di sintetizzare da sé.

In particolare tra i prodotti di quasi tutti e sei i batteri risultano alcune molecole che secondo fonti bibliografiche (McLeod *et al.*, 2004) potrebbero essere coinvolte nella virulenza:

- acyl-carrier-protein (*cofattore*);
- lipoprotein;
- colanic acid capsular biosynthesis activation protein A cleavage peptides (*polisaccaride extra cellulare*);
- N-acetil-D-glucosaminil-N-acetilmuramoil-pentapeptidil difosfoundecaprenol e N-acetil-alfa-D-glucosamina (*componenti del peptidoglicano*).

Solo *E. chaffeensis* non produce nessuno di questi ultimi due composti, non è infatti in grado di sintetizzare il peptidoglicano dato che i pathway di sintesi sono andati persi.

È interessante notare che tra le molecole prodotte alcune risultano essere esclusive delle 4 specie di Rickettsiales analizzate: sono composti lipopolisaccaridici coinvolti nel pathway del

lipide A, fosfolipidi e lipidi: 2,3,2',3'-tetrakis (3-idrossitetradecanoil)-D-glucosaminil-1,6-beta-D-glucosamina 1,4'-bisfosfato, fosfatidiletanolamina, fosfatidilglicerolo.

In generale i risultati ottenuti sono confermati in parte dai dati genomici disponibili almeno per quanto riguarda gli organismi intracellulari. La non corrispondenza con i dati di origine genomica può essere in parte ricondotta in primo luogo a errori presenti nei dati di origine: alcune molecole che i batteri sarebbero in grado di sintetizzare infatti non risultano dall'analisi della rete metabolica, altre invece risultano in uscita molecole che non potrebbero sintetizzare. È però possibile che i batteri ottengano dall'ospite i precursori, per questo motivo le molecole che non sarebbero in grado di sintetizzare *de novo* non compaiono tra le molecole in entrata: in certi casi infatti è riportato che sono presenti solo gli step finali per la sintesi; per esempio diverse forme di folato e/o precursori risultano essere importati: i batteri attraverso gli step finali di sintesi otterrebbero il tetraidrofolato, la forma attiva del coenzima; oppure hanno in entrata il pantotenato che è uno dei composti da cui parte la sintesi del coenzima A che non sono in grado di sintetizzare *de novo*.

### **2. 3. Indici di somiglianza e alberi**

La lista di molecole di ciascuna specie sono state confrontate prima con i 'solo reagenti' e 'solo prodotti' di *E. coli*: sono così state individuate le molecole che non risultano nelle liste di *E. coli* e quindi sono esclusive delle specie di Rickettsiales considerate.

In seguito queste sono state confrontate tra loro e i risultati sono riportati nelle tabelle mostrate in appendice D e E.

Le molecole che risultano essere acquisite da tutte e quattro le specie di Rickettsiales considerate rappresentano circa il 13% del totale e sono principalmente amminoacidi, nucleotidi e composti necessari per la crescita dei batteri; invece le molecole che risultano essere solo in uscita da tutte e quattro le specie di Rickettsiales considerate sono in numero ridotto rispetto al totale delle molecole solo in uscita: rappresentano circa il 10% del totale e sono per lo più molecole che costituiscono o rientrano nella biosintesi dei componenti della membrana esterna e della parete dei batteri.

Per confrontare le liste dei reagenti e dei prodotti delle quattro specie sono stati calcolati due indici di somiglianza (considerando separatamente reagenti e prodotti), l'indice di Sorensen e l'indice di Jaccard (Tab. 10, Tab. 11).

Tab. 10. Valori dell'indice di Sorensen e di Jaccard (per i reagenti)

	Indice di Sorensen					Indice di Jaccard			
	wBm	<i>E. ch</i>	<i>R. ty</i>	<i>R. pr</i>		wBm	<i>E. ch</i>	<i>R. ty</i>	<i>R. pr</i>
wBm	1	0,27	0,11	0,15	wBm	1	0,15	0,06	0,08
<i>E. ch</i>	0,27	1	0,20	0,23	<i>E. ch</i>	0,15	1	0,11	0,13
<i>R. ty</i>	0,11	0,20	1	0,26	<i>R. ty</i>	0,06	0,11	1	0,15
<i>R. pr</i>	0,15	0,23	0,26	1	<i>R. pr</i>	0,08	0,13	0,15	1

Tab. 11. Valori dell'indice di Sorensen e di Jaccard (per i prodotti)

	Indice di Sorensen					Indice di Jaccard			
	wBm	<i>E. ch</i>	<i>R. ty</i>	<i>R. pr</i>		wBm	<i>E. ch</i>	<i>R. ty</i>	<i>R. pr</i>
wBm	1	0,46	0,19	0,23	wBm	1	0,30	0,10	0,13
<i>E. ch</i>	0,46	1	0,20	0,30	<i>E. ch</i>	0,30	1	0,11	0,18
<i>R. ty</i>	0,19	0,20	1	0,52	<i>R. ty</i>	0,10	0,11	1	0,35
<i>R. pr</i>	0,23	0,30	0,52	1	<i>R. pr</i>	0,13	0,18	0,35	1

Gli indici calcolati rispecchiano le relazioni filogenetiche esistenti tra i vari taxon: in particolare la somiglianza tra *R. typhi* e *R. prowazekii* è molto simile a quella calcolata tra *E. chaffeensis* e *W. pipientis* (appartengono entrambe alla famiglia Anaplasmataceae) sia per i prodotti che per i reagenti.

Inoltre i valori di somiglianza più alti sono quelli calcolati per i prodotti, hanno un valore all'incirca doppio con entrambi gli indici.

*E. chaffeensis* e *W. pipientis* sebbene siano il primo un organismo parassita e il secondo sia primariamente un mutualista e provochino patologie molto diverse dal punto di vista delle molecole d'interazione sono simili.

Infine le varie molecole in entrata e in uscita dai vari microrganismi sono state considerate ciascuna come un carattere discreto. Considerando separatamente le molecole in entrata e quelle in uscita è stata valutata la presenza-assenza per ciascuna specie. Sono state ottenute due matrici dalle quali sono stati ricostruiti due alberi non radicati (Fig. 23).

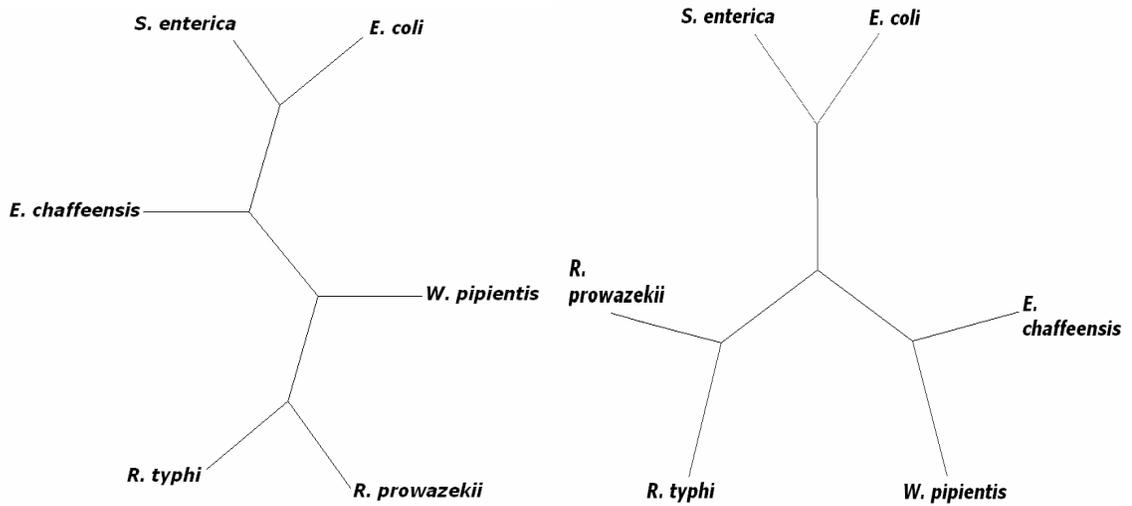


Fig. 23. Alberi filogenetici non radicati ottenuti dalle matrici di presenza-assenza delle molecole 'solo reagenti' (a sinistra) e 'solo prodotti' (destra).

Entrambi gli alberi rispecchiano i rapporti filogenetici ottenuti dalla comparazione tra le sequenze di rRNA 16S (Fig. 24): *E. coli* e *S. enterica*, due  $\gamma$ -Proteobatteri risultano più affini e sono separati dal gruppo degli  $\alpha$ -Proteobatteri. In particolare le *Rickettsia* spp. costituiscono un monophylum a conferma del fatto che sono molto simili oltre che dal punto di vista genetico (McLeod *et al.*, 2004) anche per quanto riguarda le molecole di interazione.

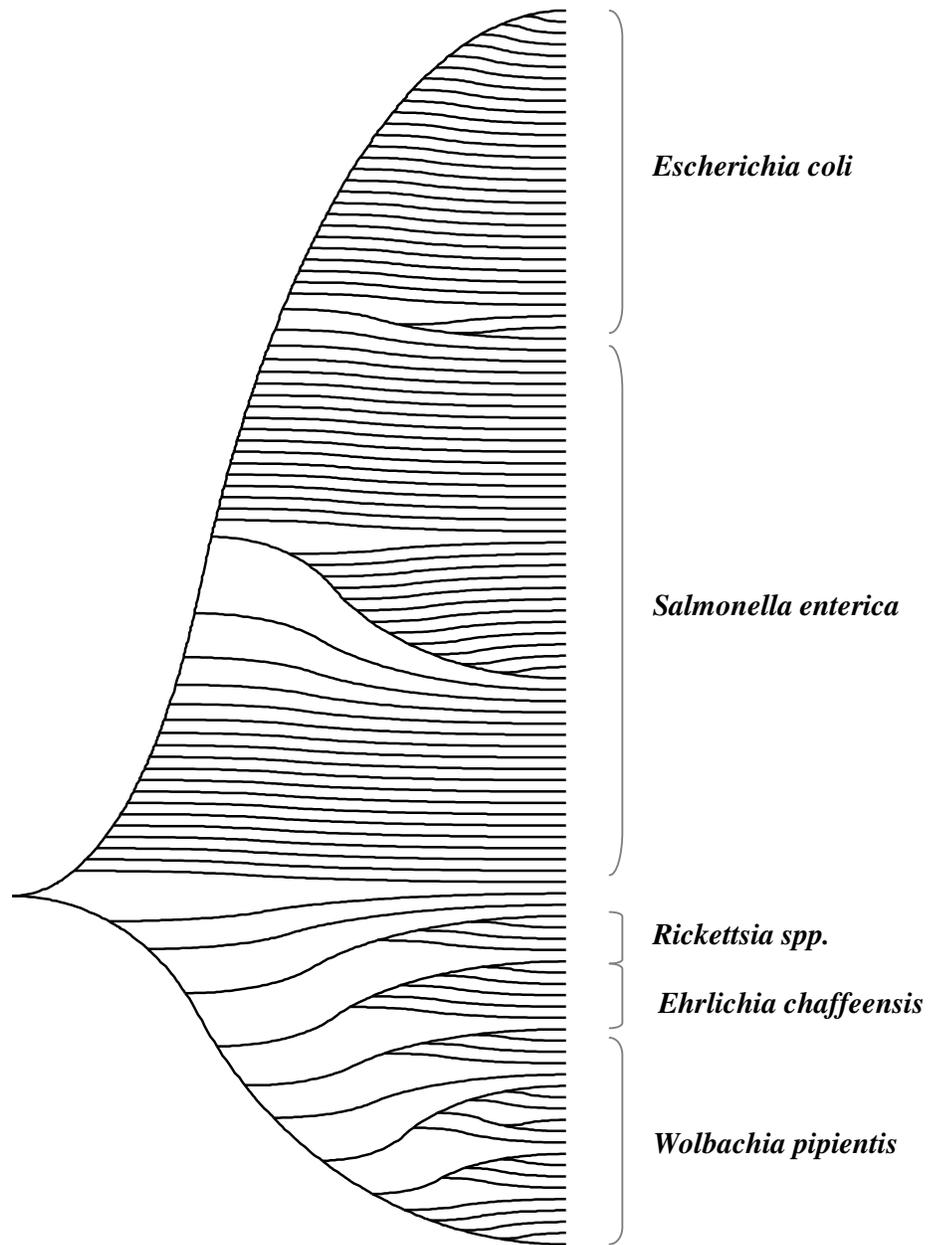


Fig. 24. Albero filogenetico (*Neighbour Joining*) basato sulle sequenze di rRNA 16S. Le ricostruzioni basate sulle sequenze disponibili dei geni *groEL* e *ftsZ* non presentano differenze significative rispetto a quella presentata.

## 5) CONCLUSIONI

In questo lavoro è stato studiato il metabolismo di diversi organismi applicando la teoria delle reti.

Dall'analisi delle reti metaboliche costruite sono emersi aspetti interessanti; tutti quanti i grafi mostrano una topologia comune: sono reti a invarianza di scala, cioè la distribuzione dei gradi dei nodi segue una legge di potenza. Tale struttura ha permesso di identificare le molecole hub che, seppur con qualche differenza tra i vari organismi, si sono rivelate essere le molecole il cui ruolo chiave nel metabolismo è già noto. Inoltre in tutte le reti è stata misurata una distanza media pari a circa 4,3 passaggi (con valori compresi tra 3,97 e 4,6 passaggi), un valore molto basso se rapportato alle dimensioni delle reti che hanno in media circa 2000 nodi che permette di definirle reti piccolo-mondo. Differenze significative sono state riscontrate nelle dimensioni delle reti: i microrganismi intracellulari, caratterizzati da un genoma ridotto e un minor numero di geni rispetto a quelli a vita libera, presentano un rete più piccola. Rapportando però il numero dei centri di reazione con il numero di geni si è osservato che, in proporzione, le specie di batteri intracellulari, caratterizzati da un contenuto genico minore, mostrano una rete metabolica più grande. Sebbene negli organismi intracellulari il processo di *genome shrinkage* abbia colpito buona parte dei geni metabolici, ciascuna specie ha mantenuto una percentuale elevata di geni in determinati ruoli funzionali riconducibili al metabolismo.

Grazie alla presenza di link direzionati è stato possibile identificare quelle molecole oggetto di scambio tra i microrganismi e l'ospite e confrontare i dati ottenuti con le informazioni rese disponibili dal sequenziamento dei genomi. Nel confronto tra *W. pipientis* e l'ospite-nematode sebbene non sia stata riscontrata una corrispondenza perfetta tra queste molecole di interazione, dovuta in parte anche all'utilizzo di *C. elegans* al posto di *B. malayi*, i dati ottenuti hanno trovato conferma nelle informazioni genomiche disponibili. Una buona parte delle molecole identificate sono infatti i metaboliti per cui il simbionte ha mantenuto il pathway e quindi è in grado di fornirli anche all'altro simbionte che invece ha perso il pathway, a confermare il tipo di relazione mutualistico che unisce i due organismi.

Inoltre sono state confrontate le molecole di scambio dei vari microrganismi patogeni: in generale le molecole che risultano essere acquisite sono amminoacidi, nucleotidi, precursori e composti necessari per la crescita dei batteri; mentre tra le molecole sintetizzate e in uscita dai batteri sono risultati principalmente lipopolisaccaridi, fosfolipidi e lipidi che costituiscono i principali componenti delle strutture extracellulari dei batteri.

Infine le molecole in entrata e in uscita sono state considerate dei caratteri discreti e la presenza/assenza nelle varie specie ha permesso di calcolare gli indici di somiglianza e di ricostruire due alberi filogenetici: entrambi rispecchiano i rapporti filogenetici ottenuti dalla comparazione tra le sequenze genetiche.

## 6) BIBLIOGRAFIA

- Albert R., Jeong H., Barabási A.-L., 1999, Diameter of the World Wide Web. *Nature* 401, pp.130-131.
- Albert R., Jeong H., Barabási A.-L., 2000, Error and attack tolerance of complex networks. *Nature* 406, pp. 378-381.
- Albert R. e Barabási A.-L., 2002, Statistical mechanics of complex networks, *Reviews Of Modern Physics* vol. 74, pp. 47-97.
- Andersson G.E., Zomorodipour A., Andersson J.O., Sichertz- Pontn T. *et al.*, 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396, pp. 133-143.
- Arita M., 2004, The metabolic world of *Escherichia coli* is not small. *PNAS* 101 (6), pp. 1543-1547.
- Baldo I., 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, pp. 179-187.
- Bandi C., Anderson T. I. C., Genchi C., Blaxter M. L., 1998 . Phylogeny of *Wolbachia* in filarial nematodes. *Proc. R. Soc. Lond B* 265, pp. 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, pp. 509-512.
- Batagelj V. e Mrvar A., 1998, Pajek-program for large network analysis. *Connections* 21, pp. 47-57.
- Bazzocchi C., Ceciliani F., Bratttig N., 2001, I simbionti batterici (*Wolbachia*) delle filarie: implicazioni per il trattamento e la patologia delle filariosi. *Ann. Ist . Super. Sanità* 37, pp. 265-273.
- Behm A., 2002, Cap. X: Metabolism, in *The Biology of Nematodes*, Edited by D. L. Lee.
- Bianconi G., e Barabási A.-L. 2001, Competition and multiscaling in evolving networks. *Europhysics Letters* 54, pp. 436-442.
- Bordenstein e Rosengaus, 2005, Discovery of a novel *Wolbachia* supergroup in isoptera. *Current Microbiology* 51 (6), pp. 393-398.
- Bourdon J. C., V De Laurenzi V., Melino G. e D Lane, 2003, p53: 25 years of research and more questions to answer. *Cell Death and Differentiation* 10, pp. 397–399.
- Bratttig N. W., Rathjens U., Ernst M., Geisinger F., Renz A., Tischendorf F. W., 2000, Lipopolysaccharide-like molecules derived from *Wolbachia* endobacteria of the filaria *Onchocerca volvulus* are candidate mediators in the sequence of inflammatory and antiinflammatory responses of human monocytes. *Microbes and Infection* 2 (10), pp. 1147-1157.
- Bratttig 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, pp. 437-445.
- Brzuszkiewicz E., Bruggemann H., Liesegang H., Emmerth M., Albermann K., Wagner C., Buchrieser C., Gottschalk G., Hacker J.and Dobrindt U., 2006, How to become a

- uropathogen: Comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains, *PNAS* 103 (34), pp. 12879–12884.
- Buchanan M., 2003, *Nexus*, Mondadori, Milano.
- Camacho J., Guiderà R., Amaral L., 2002, Analytical solution of a model for complex food webs. *Physical review E* 65, 030901 (R).
- Capra F., 1997, *La rete della vita*, Rizzoli, 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, pp. 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, pp. 1690-1698.
- Dale C. e Moran N., 2006, Molecular interactions between bacterial symbionts and their hosts. *Cell* 126, pp. 453.
- De Castro R. e Grossman J.W., 1999, Famous trails to Paul Erdős. *Mathematical Intelligencer* 21, pp. 51-63.
- De Toni A. F. e Comello L., 2005, *Prede o ragni*, Utet Libreria, Torino.
- Dorogovtsev, S. N. e Mendes, J. F. F., 2001, Language as an evolving word web. *Proceedings of The Royal Society of London, Series B, Biological Sciences* 268 (1485), pp. 2603-2606.
- 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, pp. 208-223.
- Eom Y., Lee S., Jeong H., 2006, Exploring local structural organization of metabolic networks using subgraph patterns. *Journal of Theoretical Biology* 241, pp. 823-829.
- Erdős P. e Rényi A., 1959, On random graph. *Publicationes Mathematicae* 6, pp. 290-297.
- Erdős P. e Rényi A., 1960, On the evolution of random graph. *Publications Mathematical Institute of the Hungarian Academy of Sciences* 5, pp.17-61.
- Faloutsos M., Faloutsos P., Faloutsos C., 1999, On the power-law relationships of the Internet topology. *Computer Communications Review* 29, pp. 251-262.
- Fire, A., Xu, S., Montgomery, M. K., Kostsas, S. A., Driver, S. E., and Mello, C, C, 1998, Potent and specific genetic interference by doublestranded RNA in *Caenorhabditis elegans*. *Nature* 391, pp. 806-811.
- 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.
- Fuxelius H., Darby A., Min C., Cho N., Andersson S. G. E., 2007, The genomic and metabolic diversity of *Rickettsia*. *Research in Microbiology* 158, pp. 745-753.
- Gibbons L. M., 2002, Cap. II: General organisation, in *The Biology of Nematodes*, Edited by D. L. Lee.
- Ghedini *et al.*, 2007, Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* 317, pp. 1756-1760.
- Girin e Bouletreau, 1995, Microorganism- associated variation in host infestation efficiency in a parasitoid wasp *Trichogramma bourarachae*. *Experientia* 52, pp. 398-402.

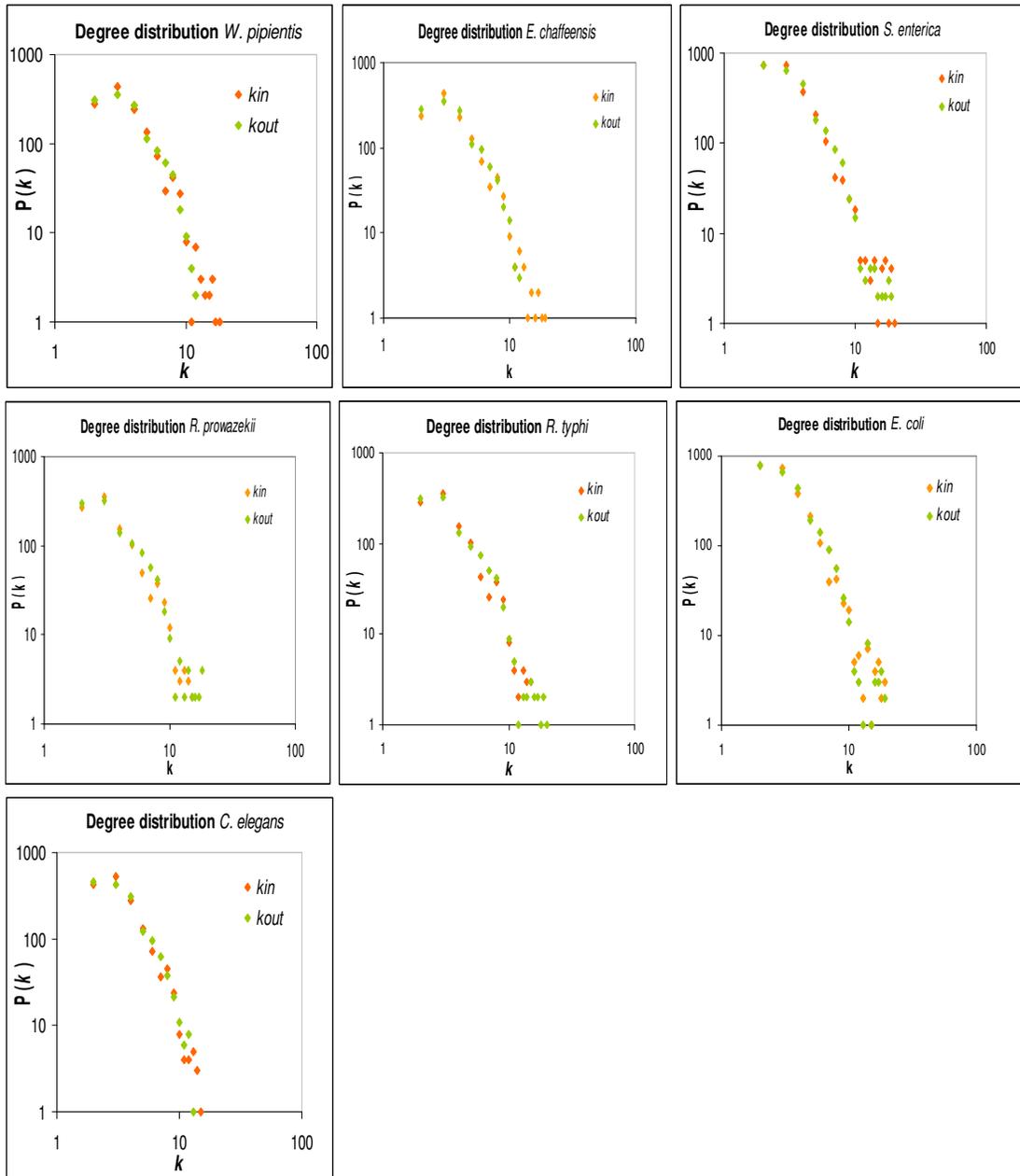
- Granovetter M., 1983, The strenght of weak ties: a network theory revisited. *Sociological Theory* 1, pp. 201-233.
- Guare J., *Six degrees of separation*, Random House, New York 1990.
- Hall T. A., 1999, BioEdit: a user-friendly biological sequence alignment editor and analysis. <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>.
- Hoffman P., 1999. *L'uomo che amava solo i numeri: la storia di Paul Erdős, un genio alla ricerca della verità matematica*. Mondadori, Milano.
- Hope I. A., 2002, Cap. V: Embryology, developmental biology and the genome, in *The biology of Nematodes*, Edited by D. L. Lee.
- 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. London ser. B.* 266, pp. 735-740.
- Jeong H., Tombor B., Albert R., Oltvai Z. N, Barabási A.-L., 2000, The large-scale organization of metabolic networks. *Nature* 407, pp. 651-654.
- Jeong H., Mason S., Barabási A.-L., 2001, Lethality and centrality in protein networks. *Nature* 411, pp. 41-42.
- 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, pp. 1341-1349.
- Karinthy F., 1929, Chains. In *Everything is different*. Budapest.
- Kumar S., Tamura K, Nei M., 2004, Mega 3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* 5, pp. 150–163.
- Koliou M., Psaroulaki A., Georgiou C., Ioannou I., Tselentis Y., Gikas A., 2007, Murine typhus in Cyprus: 21 paediatric cases. *Eur. J. Clin. Microbiol. Infect. Dis.* 26, pp. 491–493.
- Lee D. L., 2002, Cap. III: Life cycle, in *The Biology of Nematodes*, Edited by D. L. Lee.
- Lee D. L., 2002, Cap. VII: Cuticle, moulting and exsheathment, in *The Biology of Nematodes*, Edited by D. L. Lee.
- Lo N., Casiraghi M., Salati E., Bazzocchi C., Bandi, C., 2002. How many *Wolbachia* supergroups exist? *Mol Biol Evol* 19, pp. 341–346.
- Lo N., Paraskevopoulos C., Bourtzis K., O'Neill S. L., Werren J. H., Bordenstein S. R. e Bandi C., 2007, Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. *International Journal of Systematic and Evolutionary Microbiology* 57, pp. 654–657.
- 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, pp. 270-277.
- Maltsev, N., Glass, E., Sulakhe, D., Rodriguez, A., Syed, M.H., Bompada, T., Zhang, T., D'souza, M., 2006, Puma--Grid-Based highthroughput analysis of genomes and metabolic pathways. *Nucleic Acid Res.* 1; 34 (Database Issue): D369-72.
- McLeod M.P., Qin X., Karpathy 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, pp. 5842-5855.
- Moran N., 2003, Tracing the evolution of gene loss in obligate bacterial symbionts. *Current Opinion in Microbiology* 6, pp.512–518.

- Munn E. A. e Munn P., Cap. IX: Feeding and digestion, in *The Biology of Nematodes*, Edited by D. L. Lee.
- Newman M., 1999. The structure of scientific collaboration networks. *PNAS* 98 (2), pp. 404-409.
- Newman M., 2000, Models of the small world. A review. *Journal of Statistical Physics* 101, pp. 819-841.
- Newman M., Barabási A.-L., Watts D., 2006, The structure and dynamics of networks. *Princeton University Press*, Princeton.
- Page R. D. M., 1996, TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* 12, pp. 357-358.
- Pettersson M. E. e Berg O. G., 2007, Muller's ratchet in symbiont populations. *Genetica* 130, pp. 199–211.
- Price D. J. De S., 1965, Networks of scientific papers. *Science* 149, pp. 510-515.
- Przulj N., 2004, Graph theory approaches to protein interaction data analysis. *Technical Report 322/04*, Department Of Computer Science, University Of Toronto.
- Rao A. U., Carta L. K., Lesuisse E. e Hamza i., 2005, Lack of heme synthesis in a free-living eukaryote. *PNAS* 102 (12), pp. 4270–4275.
- Renesto P., Ogata H., Audic S., Jean-Michel Claverie J. M., Raoult D., Some lessons from Rickettsia genomics, 2005, *FEMS Microbiology Reviews* 29, pp. 99 – 117.
- Rousset F., Bouchon D., Pintureau B., Juchault P., Solignac M., 1992. *Wolbachia* endosymbionts responsible for various alterations of sexuality in arthropods. *Proc. R. Soc. London ser. B* 250, pp. 91-98.
- Rowley S. M., Raven R. J. e McGraw E. A., 2004, *Wolbachia pipientis* in Australian spiders. *Current Microbiology* 49 (3), pp. 208-214.
- Satorras P. R. e Vespignani A., 2001, Epidemic Spreading in scale-free networks. *Physical Review Letters* 86 (14), pp. 3200-3203.
- Satorras P. R. e Vespignani A., 2002, Immunization of complex networks. *Physical Review E* 65, 036104.
- Sironi M., Bandi C., Sacchi L., Di Sacco B., Damiani G. e Genchi C., 1995, Molecular evidence for a close relative of the arthropod endosymbiont *Wolbachia* in a filarial worm. *Molecular and Biochemical Parasitology* 74 (2), pp. 223-227.
- Solè R. V. e Montoya J. M., Complexity and fragility in ecological networks. Documento di lavoro 00-11-060, Santa Fe, Santa Fe Institute NM, 2000. Disponibile sul sito [www.santafe.edu/sfi/publications/00wplist.html](http://www.santafe.edu/sfi/publications/00wplist.html).
- Stouthamer *et al.*, 1990, Antibiotics cause parthenogenetic *Trichogramma* to revert to sex. *Proc. Natl. Acad. Sci. USA* 87, pp. 2424-27.
- Stouthamer R., Breeuwer J. A. J., Hurst G. D. D., 1999. *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Ann. Rev. Microbiol.* 53, pp. 71-102.
- Taylor M.J., Cross H. F., Bilo K., 2000. Inflammatory responses induced by the filarial nematode *Brugia malayi* are mediated by lipopolysaccharide-like activity from endosymbiotic *Wolbachia* bacteria. *J. Exp. Med.* 191, pp. 1429–1436.
- Taylor M. J., 2003, *Wolbachia* in the inflammatory pathogenesis of human filariasis. *Ann. N.Y. Acad. Sci.* 990, pp. 444–449.

- Tamas I., Klasson L., Sandstrom J., Andersson S., 2001, Mutualists and parasites: how to paint yourself into a (metabolic) corner. *FEBS Letters* 498, pp.135-139.
- The *C. elegans* Sequencing Consortium, 1998, Genome sequence of the nematode *Caenorhabditis elegans*. A platform for investigating biology, *Science* 282, pp. 2012-2018.
- Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F., Higgins D. G., 1997, The *CLUSTAL X* windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*. 25 (24) pp. 4876-4882.
- Timmons L., Fire A., 1998, Specific interference by ingested dsRNA. *Nature* 395, p. 854.
- Travers J. e Milgram S., 1969, An experimental study of the small world problem. *Sociometry* 32, pp. 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, pp. 279–286.
- Vogelstain B., Lane D., Levine A. J., 2000, Surfing the p53 network. *Nature* 408, pp. 307-310.
- Wagner A. e Fell D., 2001, The small world inside large metabolic networks. *Proceedings Of The Royal Society Of London* 268, pp.1803-1810.
- Waldrop M. M., 1995, *Complessità: uomini e idee al confine tra ordine e caos*. Instar Libri. Torino.
- Watts D. e Strogatz S.H., 1998, Collective dynamics of ‘small-world’ networks. *Nature* 393, pp. 440- 442.
- Werren J. H., Zhang W., Guo L. R., 1995. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proc. R. Soc. Lond. B* 261, pp. 55–63.
- Werren J., 1997. Biology of *Wolbachia*. *Annu. Rev. Entomol.* 42, pp. 587-609.

## 7) APPENDICI

### A) Distribuzione dei gradi dei nodi degli organismi analizzati



### B) Molecole in entrata nei microrganismi analizzati

Tab. B1 **AMMINOACIDI** in entrata

aa in entrata	<i>wBm</i>	<i>E. ch</i>	<i>R. ty</i>	<i>R. pr</i>	<i>E. co</i>	<i>S. en</i>
Corismato	+	+	+	+		
Glicina		+			+	+
L-alanina		+		+	+	+
D-alanina	+	+	+	+	+	+
L-allothreonine	+	+	+	+	+	+
L-arginina		+		+	+	
L-asparagina	+	+	+	+	+	+

L-aspartato		+			+	+
L-aspartato semialdeide	+	+	+	+	+	+
L-citrullina		+			+	+
L-cisteina		+	+	+	+	+
L-glutammato		+		+	+	+
D-glutamato	+	+	+	+		
L-glutammina		+			+	+
L-istidina		+				
L-iso-leucina		+		+	+	+
L-leucina		+		+	+	+
L-lisina		+	+	+	+	+
L-metionina		+			+	+
L-ornitina		+	+		+	+
L-fenilalanina		+		+	+	
L-prolina		+	+	+	+	
D-prolina	+				+	+
L-serina		+		+	+	
L-treonina	+	+	+	+	+	+
L-tirosina		+			+	
L-valina		+		+	+	
L-triptofano	+			+	+	
D-asparagina					+	+
D-glucosammmina					+	+
D-glutammmina					+	+
D-serina					+	+
D-treonina					+	

Tab. B2 **Composti per la sintesi POLISACCARIDI EXTACELLULARI in entrata**

Polisaccaridi extracellulari	wBm	<i>E. ch</i>	<i>R. ty</i>	<i>R. pr</i>	<i>E. co</i>	<i>S. en</i>
UDPglucose	+	+	+	+	+	+
D-glutammato	+	+	+	+		
D-alanina	+	+	+	+	+	+
L-lisina		+	+	+	+	+
UDP- <sup>N</sup> -acetylmuramoyl-pentapeptide	+		+	+	+	+

Tab. B3 **COFATTORI e VITAMINE in entrata**

Cofattori e vitamine	wBm	<i>E. ch</i>	<i>R. ty</i>	<i>R. pr</i>	<i>E. co</i>	<i>S. en</i>
Coenzima F420	+	+	+	+	+	+
Coproporfirinogeno III	+	+	+	+	+	+
Pirrolochinoline-chinone	+	+	+	+	+	+
Quinate	+	+	+	+	+	+
TMP ( <i>vit B1</i> )	+	+	+	+		
Ferrodossina (red/ox)	+	+	+	+	+	+
(R <sup>-</sup> )-pantoate	+	+	+	+	+	+
(R <sup>-</sup> )-pantothenate	+	+	+	+	+	+
FAD e FAD+	+	+		+	+	+
NAD e NAD+		+	+	+	+	+
Coenzima A				+	+	+
Piridoxina		+			+	+
Piridoxina-fosfato	+			+		
Tiamina		+			+	+
Diverse forme di folato ( <i>vit B9</i> )	+	+		+	+	+

Tab. B4 **NUCLEOTIDI in entrata**

Nucleosidi	wBm	<i>E. ch</i>	<i>R. ty</i>	<i>R. pr</i>	<i>E. co</i>	<i>S. en</i>
dGMP	+	+	+	+		
dAMP	+	+	+	+		
dCMP	+	+	+	+	+	+
dGDP	+	+		+		
dCDP	+					
dGTP		+				
dATP				+		
UDPgalactose			+			
dTDPglucose			+	+		
UMP				+	cUMP	cUMP
ADP		+				
GDP				+		
UTP				+		
CTP				+		
Adenilato						
Deossiadenosina	+	+	+	+	+	+
Guanilato						
Deossiguanosina		+			+	
Citidilato						
Deossicitidina	+	+	+	+	+	+
Timidilato						
Deossitimidina	+	+	+	+	+	+
Uridilato						
Deossicitosina					+	+
Deossiinosina					+	+
Deossiuridine					+	+

C) Molecole in uscita dai microrganismi analizzati

Tab.C1 **AMMINOACIDI in uscita**

aa in uscita	wBm	<i>E. ch</i>	<i>R. ty</i>	<i>R. pr</i>	<i>E. co</i>	<i>S. en</i>
L-omoserina		+	+	+		
L-dopa			+	+		
glutammato	+	+			+	+
D-aspartato					+	+
isocorismato					+	+

Tab.C2 **NUCLEOTIDI in uscita**

Nucleotidi in uscita	wBm	<i>E. ch</i>	<i>R. ty</i>	<i>R. pr</i>	<i>E. co</i>	<i>S. en</i>
3'AMP	+	+	+	+	+	+
3'CMP	+	+	+	+	+	+
3'GMP	+	+	+	+	+	+
3'UMP	+	+	+	+	+	+
dITP	+	+	+	+	+	+
Adenosine	+	+			+	+
Guanosina	+					
ITP	+	+		+		
pUpU	+	+	+	+	+	+
NTP	+	+	+	+	+	+
TTP		+	+	+		
UTP	+	+				

Tab. C3 **COFATTORI e VITAMINE** in uscita

Coenzimi cofattori vitamine	wBm	<i>E. ch</i>	<i>R. ty</i>	<i>R. pr</i>	<i>E. co</i>	<i>S. en</i>
protoeme	+	+	+	+	+	+
Siroeme	+	+	+	+	+	+
Menachinone	+	+	+	+	+	+
tetrahydromethanopterin	+	+	+	+	+	+
molybdopterin	+	+	+	+	+	+
coenzima F420 ridotto	+	+	+	+	+	+
FADH	+	+	+	+	+	+
diidrofolato ( <i>vit B9</i> )	+	+	+	+	+	+
ubiquinone	+	+	+	+	+	+
dihydromethanopterin	+	+	+	+	+	+
pyrroloquinoline-quinone ridotto	+	+	+	+	+	+
palmitate	+	+	+	+	+	+
biotina		+			+	+

Tab. C4 **MOLECOLE VIRULENZA** in uscita

Molecole Virulenza	wBm	<i>E. ch</i>	<i>R. ty</i>	<i>R. pr</i>	<i>E. co</i>	<i>S. en</i>
acyl-carrier-protein	+	+	+	+	+	+
Lipoprotein	+	+	+	+	+	+
colanic acid capsular biosynthesis activation protein A cleavage peptides	+	+	+	+		
(N'-acetyl-D-glucosaminyl-N'-acetylmuramoyl-pentapeptidyl)(n+1)-diphosphoundecaprenol	+		+	+		+
N'-acetyl-'alfa'-D-glucosamine	+		+	+		

D) Lista di 'solo reagenti' esclusivi delle specie di Rickettsiales analizzate

'Solo reagenti' esclusivi delle 4 specie di Rickettsiales analizzate (10) 10/ 78 = 12,82%	
3D-ribosio 1P	dGMP
4-aminobenzoato	dAMP
Beta-alanina	TMP
Corismato	IDP
D-glutamato	Isopentenil difosfato

'Solo reagenti' delle 4 specie di Rickettsiales analizzate	wBm	<i>R. pr</i>	<i>R. ty</i>	<i>E. ch</i>
(R)-3-hydroxybutanoyl	+			
N'-aminoacyl-L-prolyl-peptide	+			
1D-myo-inositol 1-P	+			
1L-myo-inositol 1-P	+			
Peptidylproline omega=180	+			
dCDP	+			
Ferricytochrome b1	+			
Oleoyl-CoA	+			
Malonate	+			
Carbamoyl phosphate	+			
[acyl-carrier protein]		+		
Myo-inositol		+		
10-formyltetrahydrofolate		+		
2-oxoglutarate		+		
5,10-methylenetetrahydrofolate		+		
Acetyl-CoA		+		

Arginate		+		
CTP		+		
D-ribose 5-P		+		
Lipoamide		+		
dATP		+		
D-fructose 6-P		+		
GDP		+		
FAD		+		
FADH <sub>2</sub>		+		
Fe <sup>2+</sup>		+		
NADH		+		
Oxaloacetate		+		
Oxidized ferredoxin		+		
Pyrophosphate		+		
Reduced acceptor		+		
Reduced ferredoxin		+		
UMP		+		
Urate		+		
UTP		+		
Succinyl-CoA		+		
Tetrahydrofolate		+		
Thiamin diphosphate		+		
tRNA		+		
NADPH			+	
Dephospho-CoA			+	
UDPgalactose			+	
Gallate			+	
AlphaDribose 1-P			+	
10-formyltetrahydromethanopterin				+
4-aminobutanoate				+
ADP				+
dGTP				+
D-ribose				+
Formaldehyde				+
Geranylgeranyl diphosphate				+
Cyclohexanol				+
CDPdiacylglycerol				+
dGDP	+	+		+
H <sub>2</sub> O		+	+	+
HCO <sub>3</sub> <sup>-</sup>	+	+	+	
Pyridoxal 5'-P	+	+	+	
`O`('3)-`sn`-phosphatidyl-L-serine		+		+
ATP		+		+
NH <sub>3</sub>		+		+
Orthophosphate		+		+
L-arginine		+		+
1,2-diacyl-`sn`-glycerol 3-P		+		+
2-oxobutanoate		+	+	
Alpha-D-glucose 1-P		+	+	
D-glucose		+	+	
D-glucose 6-P		+	+	
dTDPglucose		+	+	
Cyclic 2,3-bisphospho-D-glycerate			+	+
Glycerol			+	+

`sn`-glycerol 3-P			+	+
O <sub>2</sub>	+	+		
palmitoyl-CoA	+	+		
3-methyl-2-oxobutanoate	+			+
L-methionyl-oligopeptide	+			+
Alpha-D-glucose	+			+
Beta-D-glucose	+			+
Reduced thioredoxin	+			+

E) Lista ‘solo prodotti’ esclusivi delle specie di Rickettsiales analizzate

‘Solo prodotti’ esclusivi delle 4 specie di Rickettsiales analizzate (6) 6/58 = 10,34%
2,3,2',3'-tetrakis(3-hydroxytetradecanoyl)-D-glucosaminyl-1,6-beta-D-glucosamine 1,4'-bisphosphate
Phosphatidylethanolamine
Phosphatidylglycerol
Ethanol
L-seryl-tRNA('Ser)
Oxidized thioredoxin

‘Solo prodotti’ delle 4 specie di Rickettsiales analizzate	wBm	R. pr	R. ty	E. ch
Myo-inositol	+			
25,10-methylenetetrahydrofolate	+			
Acetone	+			
(R)-lactate	+			
L-glutamyl-tRNA('Glu)	+			
L-histidyl-tRNA('His)	+			
Thiocystine	+			
Guanosine	+			
AlphaDribose 1-P	+			
Inosine	+			
L-homocysteine	+			
Nicotinamide D-ribonucleotide	+			
Peptidylproline omega=0	+			
Succinate		+		
Malonate semialdehyde		+		
Dribose 1-P		+		
3-phospho-D-glycerate			+	
Adenosyl-cobalamin			+	
Beta-D-glucose			+	
Farnesyl diphosphate (FPP)			+	
`p`-nitrophenol				+
Formate				+
Carbamoyl phosphate				+
Glutarate				+
O <sub>2</sub>				+
Methylglyoxal				+
UTP				+
5'-methylthioadenosine				+
7,8-dihydrobiopterin				+
L-homoserine		+	+	+
`N`-acetylneuraminate		+	+	+
Catechol	+	+		+
3,4-dihydroxybenzoate	+	+		+

<b>ITP</b>	+	+		+
<b>Guanine</b>	+		+	+
<b>(S)-lactaldehyde</b>	+		+	+
<b>L-tyrosyl-tRNA('Tyr)</b>	+	+	+	
<b>Guanosine 3',5'-bis(diphosphate)</b>		+	+	
<b>CMP-3-deoxy-D-`manno`-octulosonate</b>		+	+	
<b>Butanoate</b>		+	+	
<b>IMP</b>		+	+	
<b>tRNA containing `N`('7)-methylguanine</b>		+	+	
<b>Sucrose</b>	+			+
<b>Glutathione</b>	+			+
<b>3-(3-`sn`-phosphatidyl)glycerol</b>	+			+
<b>Adenine</b>	+			+
<b>Alpha-D-glucose 6-P</b>	+			+
<b>Acetaldehyde</b>	+			+
<b>D-glycerate</b>	+			+
<b>GDPmannose</b>	+	+		
<b>CMP-`N`-acetylneuraminate</b>		+		+
<b>TTP</b>	+		+	