Supplementary Information:

Network motifs provide signatures that characterize metabolism

Erin R. Shellman[a,b], Charles F. Burant[b] and Santiago Schnell[c,b]

[a] *Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, USA.* [b] *Department of Computational Medicine & Bioinformatics, University of Michigan Medical School, Ann Arbor, USA.* [c] *Department of Molecular & Integrative Physiology, University of Michigan Medical School, Ann Arbor, USA. E-mail: schnells@umich.edu*

Supplementary material for DOI: 10.1039/C2MB25346A

Organisms Used for Analysis

In this work, the kingdom of Animalia is represented by *Homo sapiens* and *Mus musculus*. The metabolic reactions of *H. sapiens* were taken from RECON1 and the metabolic network reconstruction of *M. musculus* is largely based on RECON1 as well. Additional Eukaryotes include two species of Fungi, *Saccharomyces cerevisiae* and *Pichia pastoris*, both of which are well-known model organisms with similar life cycles and environmental requirements.

Archaea are represented by two closely related species, *Methanosarcina barkeri* and *Methanosarcina acetivorans* and a third *Halobacterium salinarum*. Both methanogenic Archaea are known for employing all three pathways of methane production, which makes them popular for use in the development of biofuels.

There are two plants under investigation, *Arabidopsis thaliana* and *Zea mays*. Similarly, there is one algae, *Chlamydomonas reinhardtii*, which belongs to the kingdom of Protista. *C. reinhardtii* is motile, and uses two flagella to propel itself. It has a light-sensitive eyespot and is a popular model organism for use in the development of alternative fuel sources.

In addition to the aforementioned organisms which comprise five of the six kingdoms of life, there are eleven species of Bacteria that span a wide range of physiological features and environments.

Clostridium thermocellum is a Gram+ anaerobe and an extremophile that lives in thermophilic environments. Many of the Bacteria in this set are interesting because of the role they play in human diseases. For instance, *Vibrio vulnificus* is a close relative of *Vibrio cholerae*, the Bacterium that causes cholera, and *Escherichia coli*, *Staphylococcus aureus*, *Helicobacter pylori* and *Mycobacterium tuberculosis* are all well-known human flora and potential pathogens, implicated in diseases of the respiratory system to the digestive tract.

Species	Kingdom	Nodes	Edges	Compartment	
A. thaliana	Plantae	1501	3411	Cytosol	[5]
		50	122	Mitochondrion	
		57	112	Peroxisome	
C. reinhardtii	Protista	660	2165	Cytosol	[3]
		25	58	Golgi	
		260	652	Mitochondrion	
		48	56	Nucleus	
C. thermocellum	Bacteria	516	1604	Cytosol	[16]
D. ethenogenes	Bacteria	501	1498	Cytosol	[1]
E. coli	Bacteria	908	2863	Cytosol	[9]
H. pylori	Bacteria	400	1194	Cytosol	[20]
H. salinarum	Archaea	526	1269	Cytosol	[11]
H. sapiens	Animalia	779	2181	Cytosol	[6]
		184	402	ER	
		234	591	Golgi	
		189	351	Lysosome	
		352	905	Mitochondrion	
		85	173	Nucleus	
		135	335	Peroxisome	
G. sulfurreducens	Bacteria	466	908	Cytosol	[14]
M. acetivorans	Archaea	697	1832	Cytosol	[13]
M. barkeri	Archaea	542	1602	Cytosol	[10]
M. musculus	Animalia	842	2399	Cytosol	[18]
		182	400	ER	
		262	643	Golgi	
		205	383	Lysosome	
		385	1019	Mitochondrion	
		85	176	Nucleus	
		140	342	Peroxisome	
M. tuberculosis	Bacteria	486	1417	Cytosol	[8]
P. pastoris	Fungi	571	1774	Cytosol	[4]
		19	22	ER	
		16	20	Golgi	
		225	576	Mitochondrion	

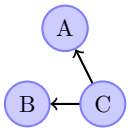
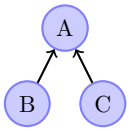
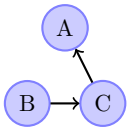
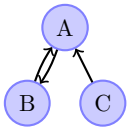
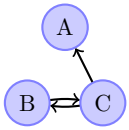
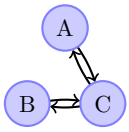
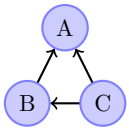
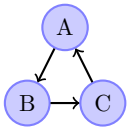
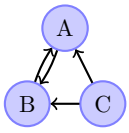
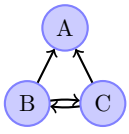
		36	62	Nucleus	
		74	161	Peroxisome	
R. etli	Bacteria	350	748	Cytosol	[15]
S. aureus	Bacteria	549	1657	Cytosol	[2]
S. cerevisiae	Fungi	528	1657	Cytosol	[7]
		15	18	ER	
		11	17	Golgi	
		214	531	Mitochondrion	
		30	45	Nucleus	
		73	186	Peroxisome	
S. typhimurium	Bacteria	852	3102	Cytosol	[19]
T. maritima	Bacteria	727	2478	Cytosol	[21]
V. vulnificus	Bacteria	831	2494	Cytosol	[12]
Z. mays	Plantae	1418	2463	Cytosol	[17]
		60	78	Mitochondrion	
		50	51	Peroxisome	

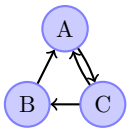
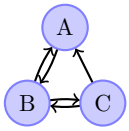
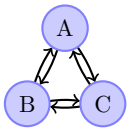
There are more published metabolic network reconstructions available than were included in the analyses. Criteria for inclusion were simply that the reconstructions were curated in SBML and were readable into the COBRA toolbox in MATLAB. These criteria insured that the reconstructions were curated using similar protocols and adequately formatted and vetted for typographical errors. Although compatability with COBRA was a requirement, neither COBRA nor MATLAB were used for analysis. Once each reconstruction was read into MATLAB, we exported relevant data as plain text files to use for motif mining. Specifically, we extracted the stoichiometric matrix, the reaction and metabolite names, a dummy variable indicating the reversibility of each reaction and the subsystem to which the reaction belonged (*e.g.* "Folate Biosynthesis," "TCA Cycle", "Salvage Pathway of ATP").

Motifs of Size Three

This work focused only on motifs of node-size three. All motifs were represented as substrate graphs. Substrate graphs represent associativity of nodes, rather than mechanistic relationships like those of a bipartite graph. Each graph type has its advantages and disadvanges. For instance, when using bipartite graphs of size 3 it is possible to generate motifs that contain no biological meaning. For example a bipartite motif might contain two nodes that represent reactions and one

that represents a metabolite, which is not a valid chemical mechanism. Similarly, because substrate graphs are associations, we cannot know the chemical mechanism from the motif structure.

Motif Number	Motif Structure	Name
1		Concurrent
2		Trapping Reactions
3		Consecutive Reactions
4		Consecutive Reactions with Reversible Step
5		Trapping Reactions with Reversible Step
6		Reversible Consecutive Reactions
7		Feed-forward Reaction
8		Closed Cycle
9		Concurrent Reaction with Exchange
10		Trapping Reaction with Exchange

11		One Way Cycle with One Reversible Step
12		One Way Cycle with Two Reversible Steps
13		Reversible Cycle

Extracting Reaction Mechanisms from Substrate Graphs

As previously mentioned, one cannot infer reaction mechanisms from substrate graphs. In order to do this as in Section 2.2, we enumerated every possible mechanism capable of yielding each of the 13 motifs. For example the first motif,  , has two possible mechanisms: It could be either C $\rightarrow$ A and C $\rightarrow$ B or C $\rightarrow$ A + B. In addition to enumerating both of these mechanisms, it is also crucial to enumerate all the combinations of reversibility. It could be the case that the correct mechanism for the first motif is C $\rightarrow$ A and C $\rightarrow$ B, but the C $\rightarrow$ B reaction is actually the reverse direction.

In order to characterize each motif, we used the stoichiometric matrices from the *E. coli*, *H. sapiens*, *M. barkeri* and *S. cerevisiae* metabolic network reconstructions. Stoichiometric matrices contain integers that denote whether a metabolite is produced, consumed or not a participant in a particular reaction. Negative integers denote consumption, positive integers denote production, and zeros denote absence. First, we generated a second stoichiometric matrix that contained the reverse mechanisms for all reversible reactions. We searched for motif mechanisms using a series of conditional tests in R. For example to find the reaction C $\rightarrow$ A, it is we used:

```
which(Stoich[paste(motif1$nodeA[i]), ] > 0 &
Stoich[paste(motif1$nodeB[i]), ] == 0 &
Stoich[paste(motif1$nodeC[i]), ] < 0)
```

The "which" function will return the stoichiometric matrix column indices for reactions where node A is being produced ($A > 0$), node B does not participate ($B == 0$) and node C is being consumed ($C < 0$). Similar conditionals were used for all other combinations of reversibility.

Results

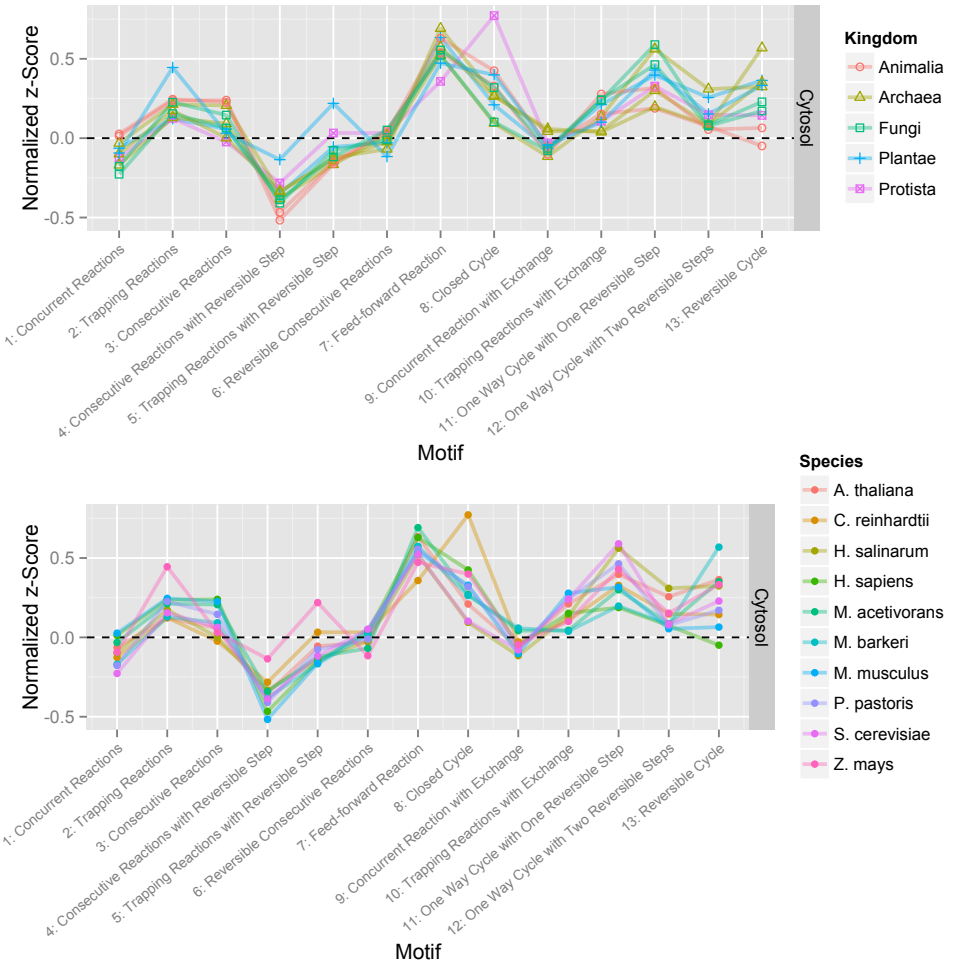


Figure S.1: Average significance profile of 3-node motifs in cytosol. The top panel shows motif distributions by kingdom and the bottom by species.

Perhaps the most striking result is the consistency of the significance profile of the cytosol across the species and kingdoms of life, which are nearly identical. This is a remarkable finding given the vast range of environments and functions of the organisms and, more practically, the

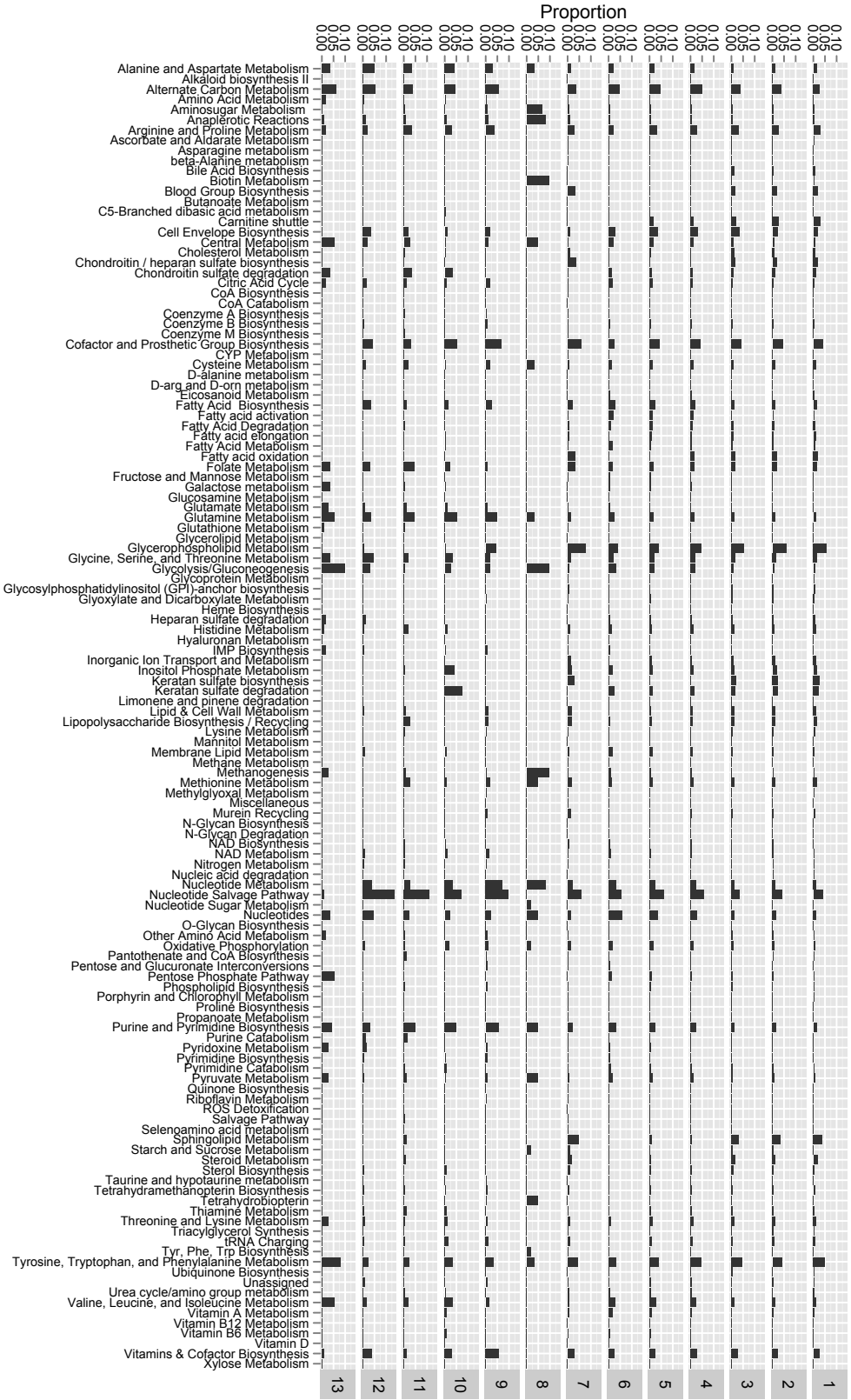


Figure S.2: Barplot showing proportion of motif participation in metabolic pathways

range of laboratories and research groups responsible for curating the reconstructions. Most of the reconstructions have differing curation protocols, naming conventions and definitions, however the emergent metabolic patterns are strikingly similar.

In addition to the overall consistency of the species, the within kingdom consistency is notable. Both species of yeast show identical enrichment/suppression patterns. There are no contrary motifs in the kingdom Animalia nor Archaea, although variation in the *extent* of enrichment is relatively high between the two species of Archaea.

Fig. S.2 contains the barplots for the proportions of motif pathway participation. In the text, this plot was reduced to only those pathways that represented at least 5% of the total for each motif.

References

[1] M Ahsanul Islam, Elizabeth A Edwards, and Radhakrishnan Mahadevan. Characterizing the metabolism of *Dehalococcoides* with a constraint-based model. *PLoS Comput. Biol.*, 6:e1000887, 2010.

[2] Scott A Becker and Bernhard Ø Palsson. Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation. *BMC Microbiol.*, 5:8, 2005.

[3] Roger L Chang, Lila Ghamsari, Ani Manichaikul, Erik F Y Hom, Santhanam Balaji, Weiqi Fu, Yun Shen, Tong Hao, Bernhard Ø Palsson, Kourosh Salehi-Ashtiani, and Jason A Papin. Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Mol. Syst. Biol.*, 7:518, 2011.

[4] Bevan KS Chung, Suresh Selvarasu, Camattari Andrea, Jimyoung Ryu, Hyeokweon Lee, Jungoh Ahn, Hongweon Lee, and Dong-Yup Lee. Genome-scale metabolic reconstruction and *in silico* analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. *Microbial Cell Factories*, 9(50):1–15, 2010.

[5] Cristiana Gomes de Oliveira Dal'Molin, Lake-Ee Quek, Robin William Palfreyman, Stevens Michael Brumbley, and Lars Keld Nielsen. AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. *Plant Physiol.*, 152(2):579–89, Feb 2010.

[6] Natalie C. Duarte, Scott A. Becker, Neema Jamshidi, Ines Thiele, Monica L. Mo, Thuy D. Vo, Rohith Srivas, and Bernhard Ø. Palsson. Global reconstruction of the human metabolic

network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA.*, 104:1777–1782, 2007.

[7] Natalie C Duarte, Markus J Herrgård, and Bernhard Ø Palsson. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.*, 14:1298–309, 2004.

[8] Xin Fang, Anders Wallqvist, and Jaques Reifman. Development and analysis of an in vivo-compatible metabolic network of *Mycobacterium tuberculosis*. *BMC Syst. Biol.*, 4:160, 2010.

[9] Adam M Feist, Christopher S Henry, Jennifer L Reed, Markus Krummenacker, Andrew R Joyce, Peter D Karp, Linda J Broadbelt, Vassily Hatzimanikatis, and Bernhard Ø Palsson. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, 3:121, 2007.

[10] Adam M Feist, Johannes C M Scholten, Bernhard Ø Palsson, Fred J Brockman, and Trey Ideker. Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Mol. Syst. Biol.*, 2:2006.0004, 2006.

[11] Orland Gonzalez, Susanne Gronau, Michaela Falb, Friedhelm Pfeiffer, Eduardo Mendoza, Ralf Zimmer, and Dieter Oesterhelt. Reconstruction, modeling & analysis of *Halobacterium salinarum* R-1 metabolism. *Mol. BioSyst.*, 4:148–159, 2008.

[12] Hyun Uk Kim, Soo Young Kim, Haeyoung Jeong, Tae Yong Kim, Jae Jong Kim, Hyon E Choy, Kyu Yang Yi, Joon Haeng Rhee, and Sang Yup Lee. Integrative genome-scale metabolic analysis of *Vibrio vulnificus* for drug targeting and discovery. *Mol. Syst. Biol.*, 18:7, 2011.

[13] Vinay Satish Kumar, James G Ferry, and Costas D Maranas. Metabolic reconstruction of the archaeon methanogen *Methanosarcina acetivorans*. *BMC Syst. Biol.*, 5:28, 2011.

[14] R. Mahadevan, D. R. Bond, J. E. Butler, A. Esteve-Nuñez, M. V. Coppi, B. O. Palsson, C. H. Schilling, and D. R. Lovley. Characterization of metabolism in the Fe(iii)-reducing organism *Geobacter sulfurreducens* by constraint-based modeling. *Appl. Environ. Microbiol.*, 72:1558–1568, 2006.

[15] Osbaldo Resendis-Antonio, Jennifer L Reed, Sergio Encarnación, Julio Collado-Vides, and Bernhard Ø Palsson. Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. *PLoS Comput. Biol.*, 3:e192, 2007.

[16] Seth B. Roberts, Christopher M. Gowen, J Paul Brooks, and Stephen S. Fong. Genome-scale metabolic analysis of *Clostridium thermocellum* for bioethanol production. *BMC Syst. Biol.*, 4:31, 2010.

[17] Rajib Saha, Patrick F. Suthers, and Costas D. Maranas. *Zea mays* *i*RS1563: A comprehensive genome-scale metabolic reconstruction of maize metabolism. *PLoS ONE*, 6(7):e21784, 07 2011.

[18] Martin I Sigurdsson, Neema Jamshidi, Eirikur Steingrimsson, Ines Thiele, and Bernhard Ø Palsson. A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Syst. Biol.*, 4: 140, 2010.

[19] Ines Thiele, Daniel Hyduke, Benjamin Steeb, Guy Fankam, Douglas Allen, Susanna Bazzani, Pep Charusanti, Feng-Chi Chen, Ronan Fleming, Chao Hsiung, Sigrid De Keersmaecker, Yu-Chieh Liao, Kathleen Marchal, Monica Mo, Emre Ozdemir, Anu Raghunathan, Jennifer Reed, Sook-Il Shin, Sara Sigurbjornsdottir, Jonas Steinmann, Suresh Sudarsan, Neil Swainston, Inge Thijs, Karsten Zengler, Bernhard Palsson, Joshua Adkins, and Dirk Bumann. A community effort towards a knowledge-base and mathematical model of the human pathogen *Salmonella typhimurium* LT2. *BMC Syst. Biol.*, 5:8, 2011.

[20] Ines Thiele, Thuy D Vo, Nathan D Price, and Bernhard Ø Palsson. Expanded metabolic reconstruction of helicobacter pylori (iit341 gsm/gpr): an in silico genome-scale characterization of single- and double-deletion mutants. *J. Bacteriol.*, 187:5818–30, 2005.

[21] Ying Zhang, Ines Thiele, Dana Weekes, Zhanwen Li, Lukasz Jaroszewski, Krzysztof Ginalski, Ashley M. Deacon, John Wooley, Scott A. Lesley, Ian A. Wilson, Bernhard Palsson, Andrei Osterman, and Adam Godzik. Three-dimensional structural view of the central metabolic network of *Thermotoga maritima*. *Science*, 325:1544–1549, 2009.