Article

Characterization *in silico* of flavonoids biosynthesis in *Theobroma cacao* L.

Arsenio Rodriguez[1, 2], Diogenes Infante[1]

[1]National Center for Agricultural Biotechnology, Institute for Advanced Studies, Apdo 17606 Parque Central Caracas 1015-A, Venezuela

[2]Graduate Program in Biology, Simon Bolívar University, Caracas, Venezuela

E-mail: dinfante@idea.gob.ve

Received 26 February 2011; *Accepted 25 March 2011*; *Published online 12 May 2011*

IAEES

Abstract

A detailed and curated map of molecular interactions taking place in the polyphenols (flavonoids) biosynthesis in *Theobroma cacao* L. is presented. The map was created using the software Cytoscape v.2.7 and the *KEGG* (Kyoto Encyclopedia of Genes and Genomes) database. The statistical parameters of the network were determined with CentiScape v.1.1 and MetaAnalyzer v.2.6.1. A preliminary theoretical map containing 1024 chemical species and 1099 chemical reactions was built, then a second map that was curated and annotated with the biological facts obtained from approximately 85 publications in *T. cacao*, with 653 chemical species and 706 chemical reactions. Structural analysis of this interaction network revealed similitude with other biological networks. The study of complex networks opens the way for creating realistic computational models of flavonoid biosynthesis metabolic pathways in cacao.

Keywords systems biology; metabolic network; graph model.

1 Introduction

Metabolism is the best characterized of all molecular interaction networks in biology. Large amounts of data related to metabolic reactions are available to date, but despite this wealth of information metabolic phenotypes still cannot be accurately predicted, both in terms of network topology and the properties of its molecular components (Sweetlove, 2008). Knowledge of plant metabolism represents challenges due to the large amount of compounds involved, the strong interaction between the genotype and the phenotype, the complex regulatory interaction implicated and the intricate interaction among them. This knowledge is important because many of them have the potential for the generation of pharmaceutical products or key secondary metabolites of commercial interest, they can enhance plant growth or defense against pest attack, and they can also improve the yield and nutritional quality of crops (nutraceuticals). Among these compounds, polyphenols are ubiquitous secondary plant metabolites that play a variety of roles in the defense mechanisms, the reproduction of plants and as antioxidant compounds (Davies, 2000). Flavonoids are water soluble and occur mostly as glycosylated compounds; they are a major subgroup of polyphenols, which assist plants in attracting pollinators and seed dispersers by providing red to blue pigmentation in flowers and fruits (Gould and Lister, 2006). These compounds also have been shown to have a significant effect on human health. They are found mainly in cocoa, tea, grapes, and olive (Vermerris and Nicholson, 2006).

The three major chemical groups reported in cocoa polyphenols are flavonols (quercetin and kampferol), flavan-3-oles (epicatechin and catechin) and anthocyanins or anthocyanidins (cyanidin), and several flavonoid compounds have been identified in cocoa: coumaric acid, amides, condensed tannins in flowers (Alemanno et. al., 2003); epicatechin (epigallocatechin gallate, epigallocatechin, gallic acid, epicatechin gallate) in leaves (Osman et. al., 2004); Quercetin, isoquercitrin, quercetin-3O-glycosides, anthocyanins (cyanidin-3-galactoside and cyanidin-3-arabinoside) in seeds (Niemenak et. al., 2006; Sánchez-Rabaneda et. al., 2003). These compounds show potent antioxidant activity *in vitro* with influence in health. A reduction in the production of atherosclerotic plaques in Kurosawa and Kusanagi-hypercholesterolemic rabbits (KHC) heritable hyperlipidemia (FH) was obtained with supplementation of cocoa polyphenols. Also, in healthy human volunteers cocoa administration protected the Low Density Lipoprotein (LDL) from oxidation (Naomi, 2002). In addition, there have been confirmed protective effects against carcinogenesis and a reduction of complications accompanying diabetes mellitus.

Metabolic pathways have been defined on the basis of their step-by-step discovery; but the advent of the enormous amount of data generated by genomic, proteomic and metabolic studies has led to the development of a mathematically based analysis of complex metabolic networks, which are able to enumerate all their pathways, take in account all requirements for cofactors and byproducts (Bailey, 1998). Thus, network-based pathways are emerging as an important tool for analysis of biological systems, because they allow the understanding of the process, since it is possible to create relatively simple models to handle complex processes (Lander, 2010), which give insights to modify the distribution of metabolic flux or to rationally design metabolic pathways for new products (Quenette and Gerard, 1993; Rodríguez and Infante, 2009).

In this paper, we report the construction of the flavonoids metabolic network model in *T. cacao* with the purpose of investigating the structural properties of the flavonoids biosynthetic network. An initial polyphenols network, consisting of 1024 nodes (chemical species) and 1099 edges (chemical reactions) was built, assuming a plant primary general model, then a second model network curated to cocoa with 653 nodes (flavonoids metabolites) and 706 reactions was obtained.

2 Methods

2.1 Network reconstruction, visualization and analysis

Two metabolic models corresponding to different levels of information were reconstructed. One is a general model that contains compounds and reactions that have been identified in the *KEGG* databases associated to the flavonoids metabolism in plants (Fig. 1). The second network derives from the first, which was manually curated to include only polyphenolic compounds and reactions that have been reported in *T. cacao* literature (approximately 85 papers) (supplementary material). The data was manually downloaded into the software Cytoscape v2.7 (Shannon et al., 2003) that was used for network assembling, visualization and determination of statistical parameters. For topological analysis, Network Analyzer v.2.6.1, CentiScape v.1.1 (Scardoni et al., 2009) and MetaAnalyzer v.2.6.1 (Assenov et al., 2008), a plug-in for Cytoscape, were used. Manual curation was used to fine-tune analysis and to solve discrepancies. Isolated compounds were not included in the network representation. As the directionality of the metabolic reactions is important, edges were directed (digraph). Two metabolites were connected by an edge if they participate as substrate and product respectively in the same reaction. Common small molecules (commodity metabolites) such as ATP, NADH, water, etc, were not removed from the network representation. A general problem in metabolic models is caused by the fact that many compounds have different protonation states depending on pH. Therefore, the balance of protons was included when available.

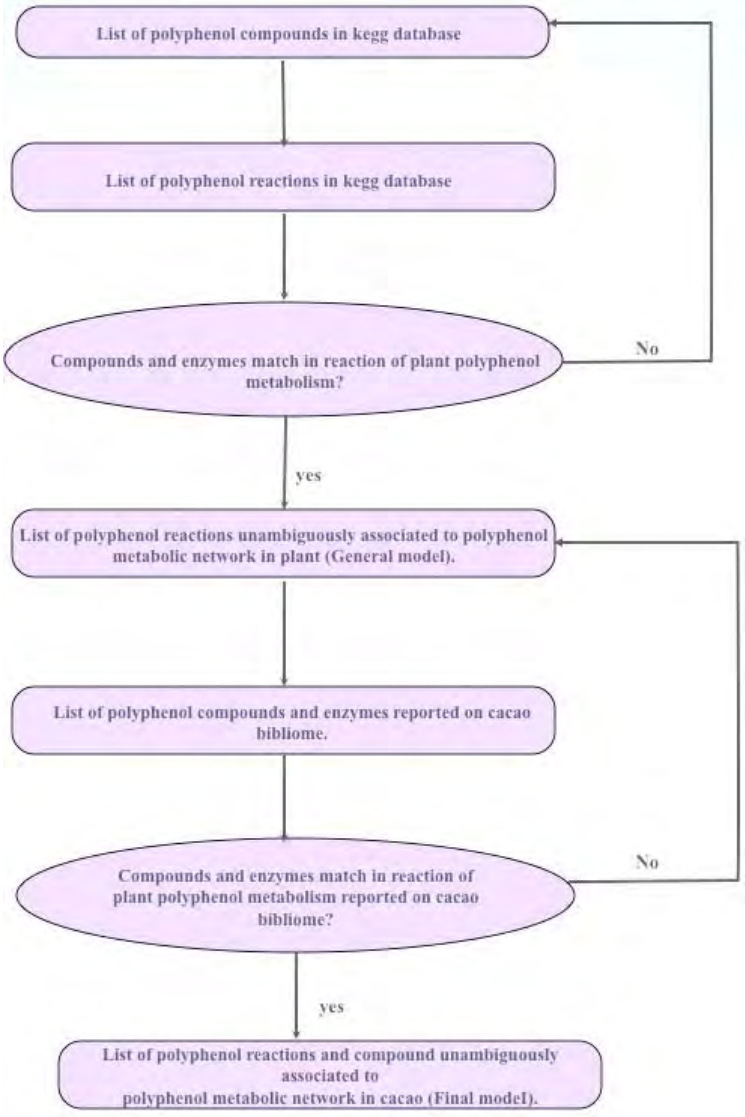

Fig. 1 Steps of the model building process.

The parameters used are explained hereafter. The number of *connected components* indicates how many disjoint sub-networks are present in the network. A *self-loop* is an edge that connects a vertex to itself. The number of *shared neighbors* between two nodes is the number of nodes that are neighbors of both of them. The *shortest path length*, also called the *distance between two nodes* is the smallest number of edges that have to be crossed to go from one node to another. The *characteristic path length* is the average distance and the *network diameter* is the largest distance between two nodes in the network. The *connectivity* of a node is the number of edges connected to it (Ying et al., 2008). The *network density* is a measure of how densely the network is populated with edges. A network that contains only isolated nodes has a density of 0, whereas a clique has a density of 1. The *network centralization* is a measure of how strongly a network is focused around central nodes; networks resembling a star have centralization close to 1, whereas decentralized networks have centralization close to 0. The *network heterogeneity* measures the variance of connectivity and reflects the

tendency of a network to contain hub nodes. The *average distance* is the average shortest path of a graph, corresponding to the summation of all shortest paths between vertex couples divided for the total number of vertex couples. The *clustering coefficient* of a given node *n* is a ratio between the number of edges between the neighbors of *n* and the maximum number of edges that could possibly exist between them. The *Betweenness centrality* $Bc_{(p)}$ of a node *p* is computed as follows:

$$Bc_{(p)} = \frac{2}{(N-1)(N-2)} \sum_{x \neq y \neq p} \frac{\varphi_{x,y(p)}}{\varphi_{x,y}},$$

where *x* and *y* are nodes in the network different from *p*, $\varphi_{x,y}$ is the number of shortest paths from *x* to *y*, $\varphi_{x,y(p)}$ is the number of shortest paths from *x* to *y* that pass through *p* and *N* is the total number of nodes in the group where *p* is connected. The *Betweenness centrality* of each node is a number between 0 and 1, which reflects the amount of control that a node exerts over the interactions between other nodes in the network. A node acting as a bridge between different communities has a high betweenness centrality, while a node that is inside a community has a low one. The *Closeness centrality* $Cc_{(q)}$ measures how close a node *q* is to others nodes in the same connected component. It is defined as follows:

$$Cc_{(q)} = \frac{N-1}{\sum_{p \neq q} L(p,q)},$$

where $L(p, q)$ is the length of the shortest path between *p* and *q*, and *N* is the total number of nodes in the connected component that *q* belongs to. The *Closeness centrality* is a measure of how fast information can spread from a given node to other reachable nodes in the network. The *centroid value* is the most complex node centrality index. It is computed by focusing the calculation on couples of nodes (v, w) and systematically counting the nodes that are closer (in terms of the shortest path) to v or w. The calculation proceeds by comparing the node distance from other nodes with the distance of all other nodes from the others, such that a high centroid value indicates that a node v is much closer to other nodes. Thus, the centroid value provides a centrality index always weighted with the values of all other nodes in the graph. Indeed, the node with the highest centroid value is also the node with the highest number of neighbors (not only first neighbors) as compared to all other nodes. In other terms, a node v with the highest centroid value is the node with the highest number of neighbors separated by the shortest path to v. The *centroid value* suggests that a specific node has a central position within a graph region characterized by a high density of interacting nodes. The "high" and "low" values are more meaningful when compared to the average centrality value of the graph G calculated by averaging the centrality values of all nodes in the graph. The eccentricity is a node centrality index. The eccentricity of a node v is calculated by computing the shortest path between the node v and all other nodes in the graph, then choosing the "longest" shortest path (let (v, K) be a path, where K is the most distant node from v). Once this path with length dist(v, K) is identified, its reciprocal is calculated (1/dist (v, K)). By doing that, an eccentricity with higher value assumes a positive meaning in terms of node proximity. Indeed, if the eccentricity of node v is high, this means that all other nodes are in proximity. In contrast, if the eccentricity is low, this means that there is at least one node (and all its neighbors) that is far form node v. Of course, this does not exclude that several other nodes are much closer to node v. Thus, eccentricity is a more meaningful parameter if it is high. Notably, "high" and "low" values are more significant when compared to the average eccentricity of the graph G, calculated by averaging the eccentricity values of all nodes in the graph.

3 Results

3.1 Metabolic models in *T. cacao*

Two models to study cocoa flavonoids metabolism were reconstructed with compounds and reactions that have been identified in the *KEGG* (Kyoto Encyclopedia of Genes and Genomes) database, which contains biochemical pathways information in plants (Kanehisa et al., 2008). The first was reconstructed as a complete hypothetical network for visualization of the general polyphenolic metabolism in plants, which include all the polyphenols pathways in plants, since all reactions and compounds to date were added. It contained consistent data and provided a high quality model with 1024 nodes and 1099 reactions. All substrates of a given reaction were connected to all products of that reaction, including common small molecules (commodity metabolites). This scenario allows generating case-based models without knowing the genome sequence of the plant, which happens for a huge number of plants.

In addition, 17 nested sub-networks: Phenylalanine, tyrosine and tryptophan biosynthesis, Phenylalanine metabolism, Tryptophan metabolism, Tyrosine metabolism, Glycolysis, Photosynthesis, Citrate cycle (TCA cycle), Pentose phosphate pathway, Pyruvate metabolism, Puromycin biosynthesis, Folate biosynthesis, Alkaloids biosynthesis, Isoquinoline alkaloid biosynthesis, Tropane, piperidine and pyridine alkaloid biosynthesis, Indole and Ipecac alkaloid biosynthesis, Acridone alkaloid biosynthesis, Biosynthesis of siderophore group nonribosomal peptides, Ubiquinone and menaquinone biosynthesis, and Benzoxazinone biosynthesis were included as nodes. The second model, a curated metabolic network model, containing the full set of known polyphenols metabolic reactions and compounds in *T. cacao* (supplementary material) was built from the first one and contained 653 compounds and 706 reactions that have been identified in the scientific literature (for *T. cacao* containing approximately 85 publications) (supplementary material).

Table 1 Network properties (metrics) of both metabolic networks.The meaning of the parameters given in this table is explained in the Methods section.

	General model	Cacao model
Number of nodes	1024	653
Number of edges	1099	706
Network density	0,002	0,003
Characteristic path length	6,16	6,10
Number of self-loops	0	0
Clustering coefficient	0,01	0,01
Connected components	1	1
Network diameter	17	15
Shortest paths	34.517 (3%)	24.663 (5%)
Average path length	6,16	6,10
Degree mean Value	2,15	2,16
Stress mean value	21.604,36	31.484,99
Eccentricity mean value	0,057	0,072
Centroid mean value	-1.003,91	-631,87
Closeness mean value	$1{,}13 \times 10^{-4}$	$2{,}11 \times 10^{-4}$

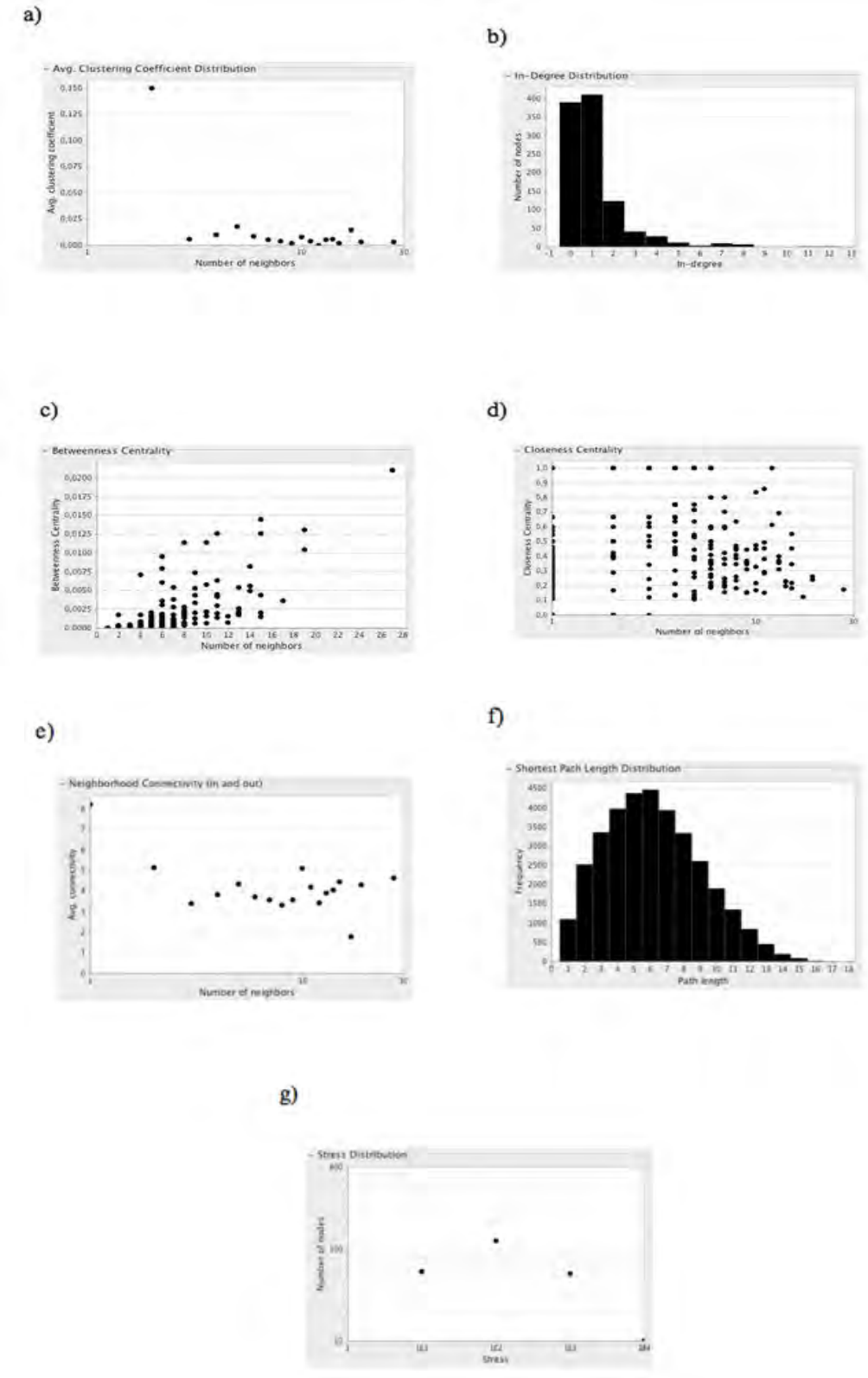

Fig. 2 Topological properties of the general model. (a) Average clustering coefficient distribution. (b) Node degree distribution (only inward). (c) Betweenness centrality. (d) Closeness centrality. (e) Neighborhood Connectivity Distribution. (f) Shortest path length distribution. (g) Stress Distribution. See Methods Section for an explanation of network parameters.

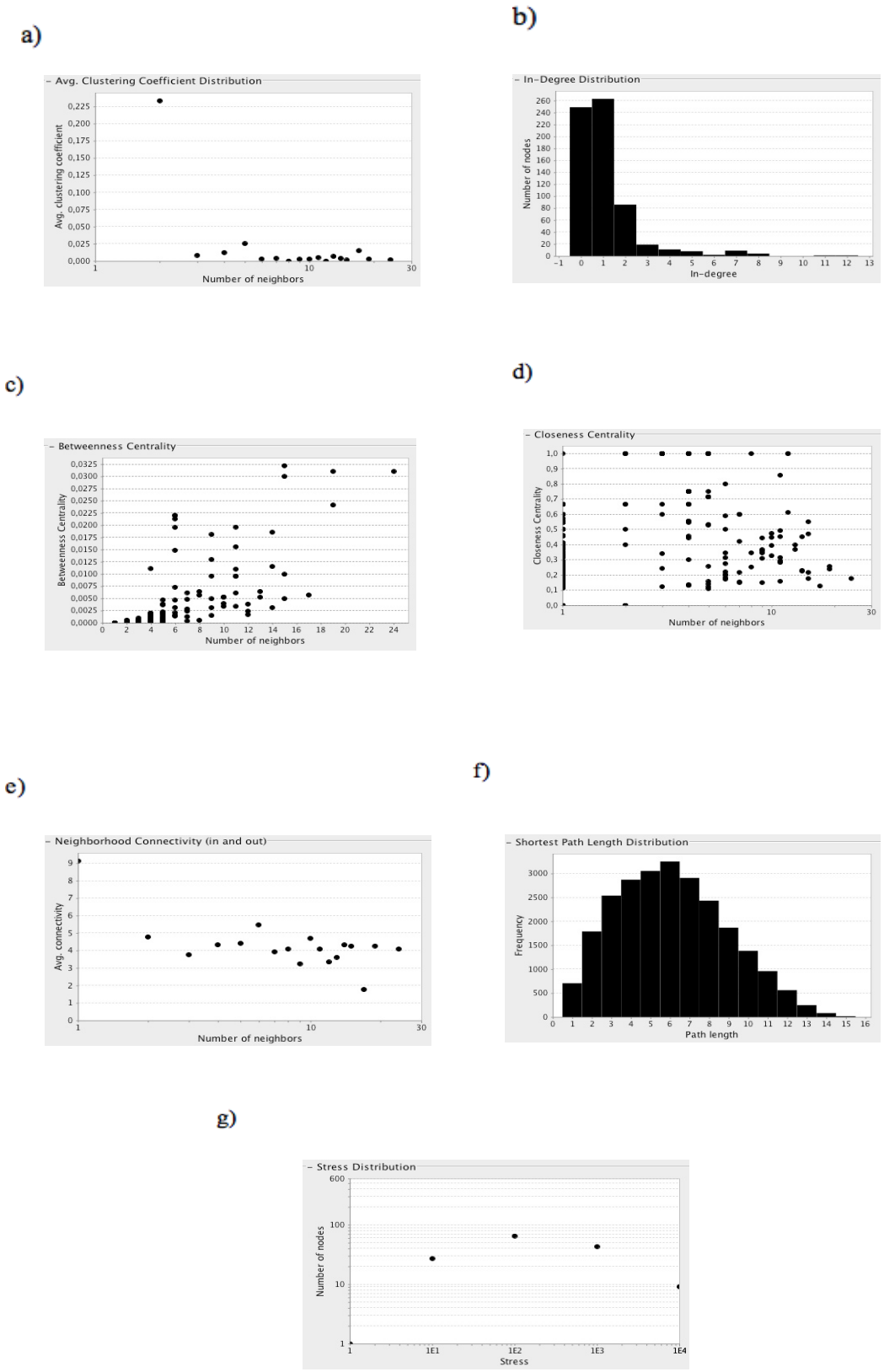

Fig. 3 Topological properties of the cocoa model. (a) Average clustering coefficient distribution. (b) Node degree distribution (only inward). (c) Betweenness centrality. (d) Closeness centrality. (e) Neighborhood Connectivity Distribution (f) Shortest path length distribution, (g) Stress Distribution. See Methods Section for an explanation of network parameters.

3.2 Network properties

We analyzed the topological properties of the reconstructed metabolic models, where metabolites were represented as nodes and reactions as edges (Fig. 2 in supplementary material). Average network parameters are shown for both networks in Table 1. The most important statistical properties for both networks are plotted in Figures 2 and 3. The average path length (APL) in our models is close to 6, which is twice the observed length in many other metabolic networks (Jeong et al., 2000), which supports the idea of lack of these small-world networks. This value becomes significantly higher when common small molecules are removed from the network (Ma and Zeng, 2003) or when an atomic representation of metabolism is adopted (Arita, 2004). APL values were higher for subnets included in our model such as Phenylalanine metabolism, pentose phosphate pathway, Glycolysis and Photosynthesis, with values ranging between 9.24 and 8.86. Phenylalanine and p-coumaroyl quinic acid metabolites had values of 8.11 and 7.62, which suggested these metabolites could have the largest mass transfer within the network (Table 2).

Table 2 Average path length for the 10 nodes.

Nodes	Average path length
Phenylalanine metabolism	9.24
Pentose phosphate pathway	8.86
Glycolysis	8.86
Photosynthesis	8.86
Quinate	8.60
Shikimate	8.54
Tyrosine metabolism	8.25
Phe	8.11
Phenylalanine, tyrosine and	7.89
NADP+	7.65
H2O	7.65
NADPH	7.65
Quinate	7.65
H+	7.65
O2	7.65
p-Coumaroyl quinic acid	7.62
O2	7.56
NADPH	7.56

The top ten hubs for the two metabolic models are listed in Table 3. p-coumaroyl-CoA remains the most highly connected molecule in both cases and nine out of ten molecules consistently appear in the top ten ranking for both models. This probably suggests a central regulatory role for this node. Interestingly, many hubs, such as ATP, NADH, phenylalanine, coenzyme A, and their derivates, serve as key compounds in the transfer of specific biochemical groups, such as phosphate groups, redox equivalents, amino groups, and acetyl groups.

Table 3 Hubs of the two metabolic networks.

General (initial) model		Final (Cacao) model	
Metabolite	Degree	Metabolite	Degree
p-Coumaroyl-CoA	27	p-Coumaroyl-CoA	24
Taxifolin (Dihydroquercetin)	20	Taxifolin (Dihydroquercetin)	20
Dihydromyricetin (Ampelopsin)	19	Dihydromyricetin (Ampelopsin)	19
Phenylalanine, tyrosine and tryptophan biosynthesis	17	Phenylalanine, tyrosine and tryptophan biosynthesis	17
Caffeoyl-CoA	16	Caffeoyl-CoA	16
Dihydrokaempferol	15	Dihydrokaempferol	15
Eriodictyol	15	Eriodictyol	15
Apigenin (4',5,7-Trihydroxyflavone)	15	Apigenin (4',5,7-Trihydroxyflavone)	14
Luteolin (3',4',5,7-Tetrahydroxyflavone)	15	Luteolin (3',4',5,7-Tetrahydroxyflavone)	15
Quercetin	15	Quercetin	15
Dihydrotricetin (Pentahydroxyflavanone)	14	Dihydrotricetin (Pentahydroxyflavanone)	14
2,7,4'-Trihydroxyisoflavanone	14	Sinapic acid	13
2-Hydroxy-2,3-dihydrogenistein	13	5-Hydroxyconiferaldehyde	13
Myricetin	13	Tricetin (5,7,3',4',5'-Pentahydroxyflavone)	12
Sinapic acid	13	5-Hydroxyconiferyl alcohol	12

Degree distribution following power-law suggesting fault tolerance properties preserved from the evolutionary point of view, as it was previously observed in others metabolic networks (Jeong et al., 2000; Wagner and Fell, 2001; Almaas, 2007) (Figures 2 and 3). The global average clustering coefficient in our models is close to 0.01, more small than the reported values of 0.20 for *E. coli*, 0.23 for *S. cerevisiae*, and 0.28 for *H. pylori* based on the same network representation (Almaas, 2007). This can be interpreted as our network lacks the small world property (Dong et al., 2001). It could be also a sign of functional convergence, because of the characteristics of the secondary metabolism of flavonoids, since the preservation of the metabolism of flavonoids for at least 400 million years between different species has been demonstrated (Ronald et al., 1994; Shirley, 1996; Rausher, 2006). Average path length is one of the three most robust measures of network topology, along with its clustering coefficient and its degree distribution. Petunidin-3-(p-coumaroyl)-rutinoside-5-glucoside, malvidin-3-(p-coumaroyl)-rutinoside-5-glucoside and cyanidin 5-O-glucoside (Table 4) have the highest polyphenols local average clustering coefficient. This suggests that these flavonoids composed a module with its own autonomy, a property that should be explored in the future.

4 Discussion and Conclusion

In this work, we introduced the first and only phenolic model of the metabolism of *T. cacao* using an alternative methodology High-quality and comprehensive models of *T. cacao* metabolism will be crucial to understand and improve cocoa from the secondary metabolism point of view.

Table 4 Clustering Coefficient for the 10 nodes.

Nodes	Average local clustering coefficient
Pyruvate metabolism	1.0
Citrate cycle (TCA cycle)	1.0
Petunidin-3-(p-coumaroyl)-rutinoside-5-glucoside	0.5
Malvidin-3-(p-coumaroyl)-rutinoside-5-glucoside	0.5
Cyanidin 5-O-glucoside	0.5
Tyrosine metabolism	0.2
Tyr	0.16
Phe	0.16
Delphinidin-3-(p-coumaroyl)-rutinoside-5-glucoside	0.16
Phenylalanine metabolism	0.15
Sinapine (Sinapoylcholine)	0.08
Cyanidin 3-O-(6-O-p-coumaroyl)glucoside	0.08
Shisonin	0.05
1-O-Sinapoyl-beta-D-glucose	0.05
Naringenin	0.03
Cyanin (Cyanidin 3,5-O-diglucoside)	0.02
Phenylalanine, tyrosine and tryptophan biosynthesis	0.01
Sinapoyl-CoA	0.01
Sinapic acid	0.01
Sinapaldehyde	0.01
Feruloyl-CoA	0.01

We reconstructed the phenolic secondary metabolism of *T. cacao* in two steps: first a general model including all the known reactions in different plants published to date was developed and second this model was curated to include only the reactions in *T. cacao*. After analysis we found that our network model is scale-free; however, it is not small world. The secondary metabolism in plants is extremely complex, highly compartmentalized, with numerous reversible reactions, many independent one way paths. It is also dependent on the stage, tissue or organ and the use of many common cofactors (ATP, NADP; NADPH). It has several different functions (defense against pest, lignin synthesis, floral pigmentation, intercellular and inter-organism communication), so these networks present a very complex topology, which must be studied in order to understand them. However, we were able to compare our model only with primary metabolism networks (i.e. glycolysis), which due to their condition are highly connected and with hub nodes, therefore small-world.

Network applications include the design of engineered biological systems, the generation of testable hypotheses regarding network structure and function and the elucidation of properties that cannot be described by simple enumeration of the individual components (such as product yield, network robustness, correlated reactions and predictions of minimal media). Recently, these properties have also been studied in genome-scale networks. To properly understand and analyze the global properties of metabolic networks, methods for rationally representing and quantitatively analyzing their structure are needed (Ma and Zeng, 2003).

The advances in genomics and molecular biology, with a huge increase in data generation, made possible contemplating and studying cellular processes as a whole in a single and coherent framework with molecular networks of metabolic processes and gene regulation. The genome-scale metabolic networks now being reconstructed from annotations of genome sequences demand new network-based definitions of pathways to facilitate the analysis of their capabilities and functions. Different networks such as protein interactions, metabolic and gene regulations are intrinsically connected and interwoven inside a cell or organism, because network information can be described and quantified (de Silva and Stumpf, 2005). Thus, network-based pathways are emerging as an important tool for the analysis of biological systems, because they allow the understanding of the process as a whole, since it is possible to create relatively simple models to handle complex processes (Lander, 2010). After the first model reconstructions were published for *E. coli* and *S. cerevisiae* (Edwards and Palsson, 2000; Förster et al., 2003), the number of reconstructions has been growing in recent years, covering many microorganisms, animals, plants and humans. A comprehensive description of those reconstructions has been presented by Feist and Palsson (2008). The applications include network property analysis, metabolic engineering, biological discovery, phenotypic assessment and evolutionary analysis. Metabolic engineering is of particular significance in plants and offers promising perspectives to improve production yields, enhancing the nutritional value of crops, generating valuable molecules for pharmacology and energy production. But for most plants and microorganisms whose genomes have not yet been sequenced, alternative methods and strategies should be developed to achieve the understanding of their physiology and metabolism through the use of general network models. The secondary metabolism networks of plants have higher complexity than those described to date in other living organisms; so, the reconstruction of such models is therefore both relevant and timely.

References

Alemanno L, Ramos T, Gargadenec A, et al. 2003. Localization and identification of phenolic compounds in *Theobroma cacao* L. somatic embryogenesis. Annals of Botany, 92(4): 613-623

Almaas E. 2007. Biological impacts and context of network biology. Journal of Experimental Biology, 210: 1548-1558

Arita M. 2004. The metabolic world of *Escherichia coli* is not small. Proceedings of the National Academy of Sciences of the United States of America, 101(6): 1543-1547

Assenov Y, Ramírez F, Schelhorn SE, et al. 2008. Computing topological parameters of biological networks. Bioinformatics, 24: 282-284

Bailey JE. 1998. Mathematical modeling and analysis in biochemical engineering: Past accomplishments and future opportunities. Biotechnology Progress, 14: 8–20

Davies KM. 2000. Plant colour and fragrance. In: Metabolic Engineering of Plant Secondary Metabolism (Verpoorte R, Alfermann AW eds). Kluwer Academic Publishers, 127-163

de Silva E, Stumpf MP. 2005. Complex networks and simple models in biology. Journal of the Royal Society Interface, 2(5): 419-430

Dong X, Braun EL, Grotewold E. 2001. Functional conservation of plant secondary metabolic enzymes revealed by complementation of Arabidopsis flavonoid mutants with maize genes. Plant Physiology, 127(1): 46-57

Edwards JS, Palsson BØ. 2000. Robustness analysis of the *Escherichia coli* metabolic network. Biotechnology Progress, 16(6): 927-939

Feist AM, Palsson BØ. 2008. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. Nature Biotechnology, 26: 659-667

Förster J, Famili I, Fu P, et al. 2003. Metabolic Network Genome-Scale Reconstruction of the *Saccharomyces cerevisiae*. Genome Research, 13: 244-253

Gould KS, Lister C. 2006. Flavonoid functions in plants. In: Flavonoids: Chemistry, Biochemistry and Applications (Andersen ØM, Markham KR eds). CRC Press, Boca Raton, 397–411

Jeong H, Tombor B, Albert R, et al. 2000. The large-scale organization of metabolic networks. Nature, 407: 651-654

Kanehisa M, Araki M, Goto S, et al. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Research, 36: D480-484

Lander A. 2010. The edges of understanding. BMC Biology, 8(40): 1-4

Ma H, Zeng AP. 2003. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics, 19(2): 270-277

Naomi O. 2002. Potential health benefits and bioavailability of polyphenols in cacao. Scientific Reports of Meiji Seika Kaisha, 41: 69-82

Niemenak N, Rohsius C, Elwers S, et al. 2006. Comparative study of different cocoa (*Theobroma cacao* L.) clones in terms of their phenolics and anthocyanins contents. Journal of Food Composition and Analysis, 19(6-7): 612-619

Osman H, Nasarudin R, Lee SL. 2004. Extracts of cocoa (*Theobroma cacao* L.) leaves and their antioxidation potential. Food Chemistry, 86(1): 41-46

Quenette PY, Gerard JF. 1993, Whywhy biologists do not think like newtonian physicist. Oikos, 68(2): 361-363

Rausher M. 2006. The evolution of Flavonoids and their genes. In: The Science of Flavonoids (Grotewold E ed). Springer, 175-211

Rodríguez A, Infante D. 2009. Network models in the study of metabolism. Electronic Journal of Biotechnology, 12(3) : 1-19

Ronald E, Koes RE, Quattrocchio F, et al. 1994. The flavonoid biosynthetic pathway in plants: function and evolution. BioEssays, 16(2): 123–132

Sánchez-Rabaneda F, Jáuregui O, Casals I, et al. 2003. Liquid chromatographic/electrospray ionization tandem mass spectrometric study of the phenolic composition of cocoa (*Theobroma cacao*). Journal of Mass Spectrometry, 38(1): 35-42

Scardoni G, Petterlini M, Laudanna C. 2009. Analyzing biological network parameters with CentiScaPe. Bioinformatics, 25(21): 2857–2859

Shannon P, Markiel A, Ozier O, et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research, 13: 2498-504

Shirley BW. 1996. Flavonoid biosynthesis: ‘new’ functions for an ‘old’ pathway. Trends in Plant Science, 1(11): 377-382

Sweetlove LJ, Fell DA, Fernie AR. 2008. Getting to grips with the plant metabolic network. Biochemical Journal, 409: 27-41

Vermerris W, Nicholson R. 2006. Phenolic compound and their effects on human health. In: Phenolic Compound Biochemistry. Springer, 235-255

Wagner A, Fell D. 2001. The small world inside large metabolic networks. Proceedings of the Royal Society B, 268: 1803-1810

Ying L, Cao HD; Shan XM, et al. 2008. An estimation formula for the average path length of scale-free networks. Chinese Physics B, 17(7): 2327