New Phytologist

Research

Metabolic and evolutionary costs of herbivory defense: systems biology of glucosinolate synthesis

Michaël Bekaert[1,2], Patrick P. Edger[3], Corey M. Hudson[4], J. Chris Pires[3,4] and Gavin C. Conant[2,4]

[1]Institute of Aquaculture, University of Stirling, Stirling, FK9 4LA, UK; [2]Division of Animal Sciences, University of Missouri, Columbia, MO, 65211, USA; [3]Division of Biological Sciences, University of Missouri, Columbia, MO, 65211, USA; [4]Informatics Institute, University of Missouri, Columbia, MO, 65211, USA

Author for correspondence:
Michaël Bekaert
Tel: +44 (0)1786 467870
Email: michael.bekaert@stir.ac.uk

Received: *13 June 2012*
Accepted: *13 July 2012*

New Phytologist (2012)
doi: 10.1111/j.1469-8137.2012.04302.x

Key words: adaptation, *Arabidopsis thaliana*, costly traits, flux balance analysis, metabolic network.

Summary

- Here, we describe our updated mathematical model of *Arabidopsis thaliana* Columbia metabolism, which adds the glucosinolates, an important group of secondary metabolites, to the reactions of primary metabolism. In so doing, we also describe the evolutionary origins of the enzymes involved in glucosinolate synthesis. We use this model to address a long-standing question in plant evolutionary biology: whether or not apparently defensive compounds such as glucosinolates are metabolically costly to produce.
- We use flux balance analysis to estimate the flux through every metabolic reaction in the model both when glucosinolates are synthesized and when they are absent. As a result, we can compare the metabolic costs of cell synthesis with and without these compounds, as well as inferring which reactions have their flux altered by glucosinolate synthesis.
- We find that glucosinolate production can increase photosynthetic requirements by at least 15% and that this cost is specific to the suite of glucosinolates found in *A. thaliana*, with other combinations of glucosinolates being even more costly.
- These observations suggest that glucosinolates have evolved, and indeed likely continue to evolve, for herbivory defense, since only this interpretation explains the maintenance of such costly traits.

Introduction

The quantitative understanding of evolution is rooted in population genetics (Lynch, 2007). However, one price of this mathematical rigor is a certain degree of abstraction, with genetic differences between individuals described primarily by their population frequencies and fitness effects (Hartl & Clark, 1997). Although both parameters are (in principle) measurable, neither gives direct insight into the molecular nature of the genetic differences. Recently, researchers have begun to bridge this gap between cellular biology and evolutionary genetics by linking cellular models to those of population genetics (Elena & Lenski, 2003). For instance, Dekel & Alon (2005) developed a model of the cost of gene expression in *Escherichia coli* and validated that model with experimental measurements of fitness. Their example is illustrative of an important question: when does natural selection purge 'costly' traits? In microorganisms at least, excess gene expression is costly enough to make mutations such as gene duplications potentially selectively detrimental (Wagner, 2005; Stoebel *et al.*, 2008). However, addressing these same types of questions in multicellular organisms is complicated by several factors:

- Longer generation times limit the power of experimental evolution experiments.
- Small effective population sizes raise the question of which traits are truly adaptive (Lynch, 2007).
- The genetic and phenotypic complexity of these organisms makes mapping the phenotypic traits back to their genetic precursors even more challenging.

In spite of these difficulties, plant defenses have long been a system for studying the evolution of traits that are simultaneously costly and beneficial (Rausher, 2001). A good example is the glucosinolates, chemicals that many eudicots use to protect against insect herbivory and oviposition (Kliebenstein *et al.*, 2005; Windsor *et al.*, 2005; Burow *et al.*, 2006; Halkier & Gershenzon, 2006; Beekwilder *et al.*, 2008). In this model, researchers first asked whether glucosinolates protected against herbivory and whether that protection was costly. Working in *Arabidopsis thaliana* (thale cress), Mauricio & Rausher (1997) found an association between glucosinolate production and increased fitness (seed production) when predators were present. However, the cost of this production was a fitness defect when predators were absent (Mauricio, 1998). One likely source of this defect is allocation costs, where the use of resources for defense reduces those available for other processes. In order to assess whether such allocation costs exist, it is first necessary to understand the chemical nature and evolutionary origins of the glucosinolates.

The order Brassicales shares four major classes of glucosinolates (i.e. aromatic, indole, aliphatic and methionine-derived aliphatic; Rodman *et al.*, 1998; Hall *et al.*, 2004; Ronse De Craene & Haston, 2006): the distribution of the ability to synthesize the

various glucosinolate compounds generally follows the phylogeny in Brassicales, with some interesting exceptions (Mithen *et al.*, 2010). These classes are differentiated by both their amino acid precursors and their final structure (Fig. 1). Glucosinolates derived from phenylalanine or tyrosine are termed aromatic, while those from alanine, valine, leucine and isoleucine are termed aliphatic. Both classes have been described outside the Brassicales (Rodman *et al.*, 1998; Fahey *et al.*, 2001). The indole (synthesized from tryptophan) and methionine-derived aliphatic glucosinolates are unique to the Brassicales, Indeed, the methionine-derived compounds are unique to the Brassicaceae, or mustard, family, and these are what give mustard its characteristic sharpness. Glucosinolates generally appear to be constitutively synthesized at rather low concentrations, but their synthesis is induced by herbivory through jasmonate and other signaling pathways (Textor & Gershenzon, 2009).

The distribution of the genes responsible for glucosinolate production largely mirrors their distribution, and they seem to owe much of their diversity to whole-genome duplication (WGD). The *A. thaliana* genome contains the remnants of three WGDs termed At-α, At-β and At-γ (Simillion *et al.*, 2002; Blanc *et al.*, 2003; Bowers *et al.*, 2003). The two most recent of these events, At-α and At-β, are unique to the Brassicales (Ming *et al.*, 2008; Soltis *et al.*, 2009; Schranz *et al.*, 2011). Recent evidence, which includes syntenic data, suggests that the origin of the indole and methionine-derived pathways arose post-At-β and around the time of At-α, respectively (Schranz *et al.*, 2011). The duplicated genes retained after the α event include both regulators of glucosinolate production (e.g. Myb transcription factors) and enzymes responsible for their biosynthesis, such as the enzymes of the branched-chain amino acid pathway and many downstream enzymes (e.g. CYP79s and CYP83s; Fig. 1). For example, a paralogous group of transcription factors differentially regulate the major glucosinolate classes. The MYB28/MYB29/MYB76 set regulates aliphatic glucosinolate production (Fig. 1b), while MYB34/MYB51/MYB122 regulate indole glucosinolates (Gigolashvili *et al.*, 2007, 2008; Lyons *et al.*, 2008). Importantly, the enzymes responsible for the synthesis of most glucosinolates (Fig. 1a) have been described and also show a history of duplication (Halkier & Gershenzon, 2006).

In this work, we explore the metabolic allocation costs associated with glucosinolate production. Starting from the genome-scale metabolic model of *A. thaliana* of de Oliveira Dal'Molin *et al.* (2010), we have added reactions describing glucosinolate production, derived from an extensive literature survey. This systems approach to evolution (Loewe, 2009) allows us to understand the global metabolic effects of glucosinolate production. Although the complexity of multicellular eukaryotes (Shlomi *et al.*, 2008) means that we cannot yet use our model to directly calculate fitness as is done in microorganisms (Edwards & Palsson, 2000), we are able to show nontrivial metabolic tradeoffs associated with this production (Agrawal & Fishbein, 2006; Agrawal, 2011; Agrawal & Heil, 2012). This result supports the idea that energetic limitations, in addition to other potential costs of glucosinolate production such

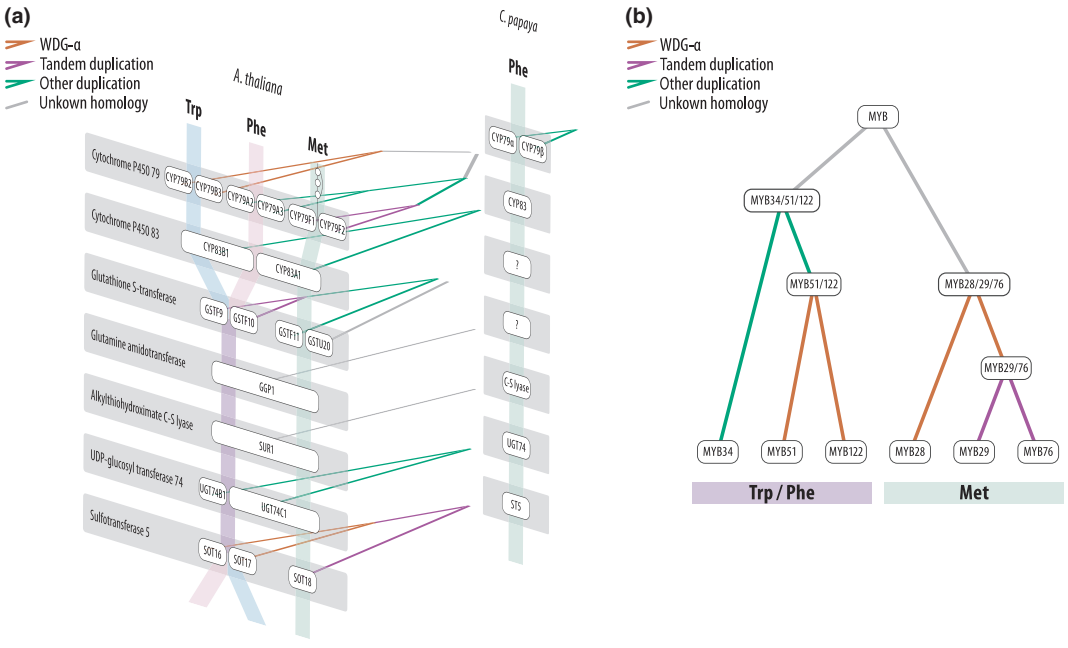

Fig. 1 Illustration of the gene duplication history of the glucosinolate pathways in *Arabidopsis thaliana*. (a) Core pathway enzymes and their evolutionary histories, including the origin of each duplication, if known. The equivalent genes in *Carica papaya* are provided for comparison. Note that, based on synthetic data, *C. papaya* lacks both At-α and At-β events. Phylogenies indicate the divergence of homologous enzymes in *A. thaliana* since the split with *C. papaya*. However, for the glutathione S-transferase and glutamine amidotransferase steps, homologous genes are not known in *C. papaya*. Multifunctional enzymes are illustrated as crossing pathways (e.g. SUR1). (b) *A. thaliana* transcriptional regulators. Methionine pathway: MYB28 (AT5G61420), MYB29 (AT5G07690) and MYB76 (AT5G07700). L-phenylalanine/L-tryptophan pathways: MYB34 (AT5G60890) MYB51 (AT1G18570) and MYB122 (AT1G74080). A larger, neighbor-joining phylogeny of the MYB family has been published and is largely congruent with these data (Sønderby *et al.*, 2007). However, that work did not describe the relationships of these genes in terms of the *Arabidopsis* WGD events.

as toxic side-effects and elemental stresses (Strauss *et al.*, 2002; Todesco *et al.*, 2010), contribute to the fitness costs of glucosinolates. In addition, it points to coevolution between plants and their predators.

Description

Data collection

The main biochemical reactions of the *A. thaliana* glucosinolate pathways were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG; release 58; Kanehisa *et al.*, 2008) and the Plant Metabolic Network (PMN; AraCyc release 8.0; Rhee *et al.*, 2006). Our manual reconstruction retained all common reaction attributes from KEGG and PMN, including unique reaction names and reaction reversibility. Chemical compound names were validated using the Chemical Identifier Resolver (http://cactus.nci.nih.gov/chemical/structure). The details of each reaction (name, EC number, subtract nature and stoichiometry) and, most importantly, the genes involved were then retrieved from the literature (see references in the Supporting Information, Table S1). These reactions were integrated into the core *A. thaliana* primary metabolic network, version 1.0.1 of which was obtained from de Oliveira Dal'Molin *et al.* (2010; AraGEM). This network includes a list of metabolites and their respective cellular compartments as well as the biochemical equation for each reaction.

Manual curation

Each metabolic reaction is defined as a node in our reaction-centric network. Edges between these nodes are defined by shared metabolites as previously described (Bekaert *et al.*, 2011). Networks were visualized with Gephi v0.7 beta (Bastian *et al.*, 2009).

Several of the newly added reactions are unbalanced, mainly because the biochemical conditions have not been completely studied. In one case (AOP2), PMN reports the production of methanesulfonate, a metabolite unknown in the plant kingdom. For the purposes of our *network* analysis, we removed this metabolite, creating an unbalanced reaction. As an aside, we note that AOP2 is known to be nonfunctional in the sequenced Columbia strain of *A. thaliana* (Kliebenstein *et al.*, 2001); in the following, we use gene expression data from that strain to limit the reactions allowed to carry flux in our model. This limitation enforces the absence of the AOP2 gene product from Columbia's glucosinolate pathways. Similarly, adding a reaction converting methanesulfonate to methanethiol did not change our flux estimates (data not shown). A second issue that may represent an important biological reality is overlapping specificities among enzymes. Thus, SOT16/17/18 have overlapping substrate specificity (Klein *et al.*, 2006). In particular, SOT16 seems to be mainly involved in L-homophenylalanine and L-tryptophan pathways and, to a lesser extent, in the L-phenylalanine pathway; SOT17 appears to be optimized for L-phenylalanine use, but also metabolizes some L-methionine derivatives. Finally, SOT18 is L-methionine-specific. For clarity, we have used only the apparent primary specificity of each enzyme in our network.

Note also that we have previously added the mitochondrial succinate/fumarate antiport from Catoni *et al.* (2003) to AraGEM in order to allow DNA/RNA synthesis (Bekaert *et al.*, 2011). The AraGEM reaction R03600 was corrected for substrate stoichiometries and the use of H^+ as a substrate. We also had to add a reaction to recycle the adenosine-3′,5′-bisphosphate created during the glucosinolate biosynthesis (KEGG R00188; Mugford *et al.*, 2009). The resulting model, AraGEM version 1.2 (Fig. 2b), is available in System Biology Markup Language (SBML) format (Notes S1).

Coexpression lookup

Gene expression data were downloaded from AtGenExpress's global stress expression data set for each gene in the glucosinolate network (Kilian *et al.*, 2007). We used R to calculate Pearson's correlation coefficient for each pairwise combination of expression profiles. Gene pairs with correlation coefficients higher than $r = 0.5$ were connected as edges in a gene coexpression network, using R's igraph network library (Csardi & Nepusz, 2006). Network coexpression clusters were assigned based on the 'Walktrap community' finding algorithm, which identifies densely connected subnetworks based on edge-weight (Pons & Latapy, 2005). In this case, edge-weight corresponds to the coexpression correlation between genes.

Duplication history

Retained duplicates from whole-genome duplications (At-α and At-β), tandem duplications in the *A. thaliana* (TAIR v10 unmasked) genome and orthologous loci in the *C. papaya* (v0.5 unmasked) genome were identified using the CoGe – GEvo tool (Lyons & Freeling, 2008, 2008) with default (B)LastZ parameters and a syntenic block size cutoff of minimally 10 shared colocalized loci. Retained duplicates from whole-genome duplications were confirmed from the list given by Bowers *et al.* (2003) as modified by Thomas *et al.* (2006).

Flux balance analysis

To perform flux balance analysis (FBA; Orth *et al.*, 2010) on the networks, we used both the Systems Biology Research Tool v2.0.0 (Wright & Wagner, 2008) and our own implementation of FBA. We computed the maximal biomass production rate under photosynthetic conditions (the import of a fixed number of photons allowed and sugar imports forbidden) with or without glucosinolates (inclusion of the glucosinuate products in the biomass reaction; Brown *et al.*, 2003). The optimum flux distribution is here defined as the flux distribution that maximizes biomass production for a given number of input photons. The rate of biomass synthesis was chosen based on the original version of AraGEM (de Oliveira Dal'Molin *et al.*, 2010). This biomass composition was then updated with results from Bekaert *et al.* (2011) defining the per-cell DNA and RNA concentrations, amino acid proportions from Schuster *et al.*, 2006 and glucosinolate masses (Table S2) taken from Brown *et al.* (2003).

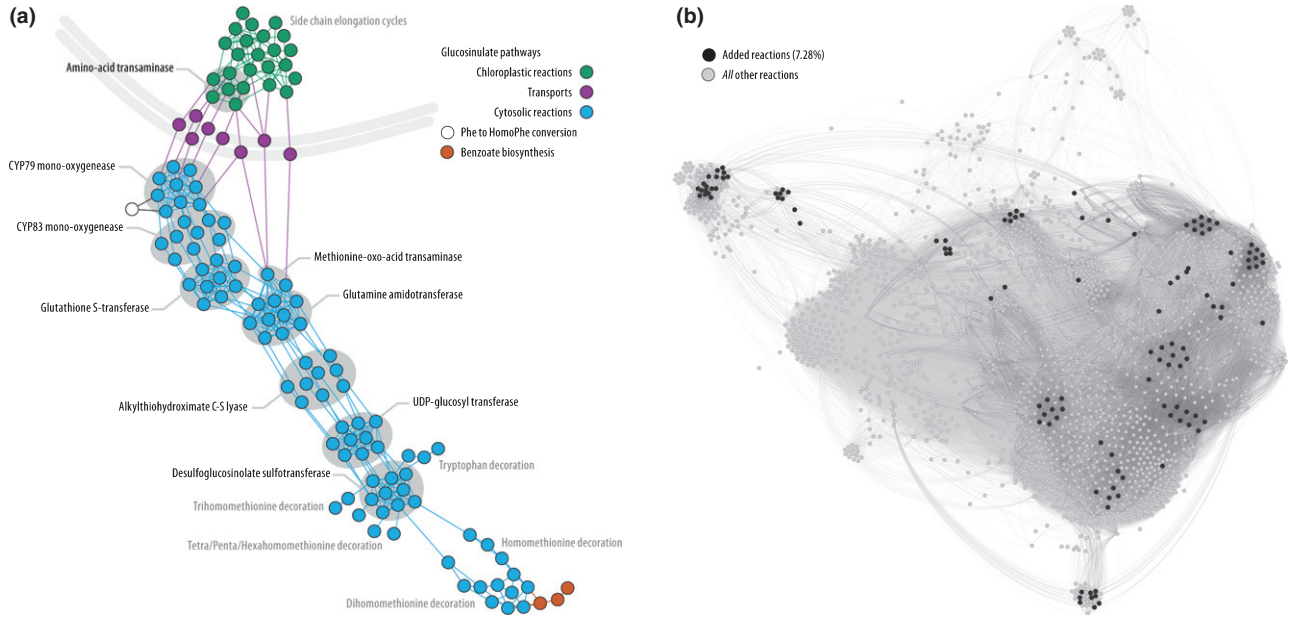

Fig. 2 Glucosinolate pathways. (a) Network illustrating the glucosinolate synthesis pathways (phloem cells), including all the chemical species of glucosinolate (see Table S2). The chain elongation cycles of the aliphatic glucosinolates (methionine-derived) occur in the chloroplast, while all other reactions occur in the cytosol. The cytochromes P79 and P83 are on the cytosolic face of the endoplasmic reticulum. The exact process for converting phenylalanine to homophenylalanine (white node) is unknown. (b) Implementation of the pathways in AraGEM. The added glucosinolate synthesis pathway accounts for 7.28% of the reactions.

We collected the expression status of all available genes in the network from an experiment measuring diurnal gene expression in *A. thaliana* Col-0 rosette leaves (sample from NCBI Gene Expression Omnibus database accession GSM77061; Blasing *et al.*, 2005). For every reaction where the catalyzing enzyme is not expressed under certain conditions, we constrained the corresponding flux to zero (reactions with no annotated genes assigned were unaltered). The remainder of the network was re-optimized and the effect of the glucosinolate production was evaluated.

The significance of the energy cost (photon uptake) was evaluated by comparing it with the distributions of 1000 permutated datasets, where the values for the 13 glucosinolates present in the biomass reaction were randomly reassigned and the flux though the network recalculated.

Results

Construction and implementation of the pathways

We adopted a reaction-centered view of *A. thaliana* metabolism, where each metabolic gene is assigned to one or more reactions. The *Arabidopsis* glucosinolate pathway (AraGLS) was reconstructed from the *A. thaliana* gene and reaction database available from the KEGG (release 58; Kanehisa *et al.*, 2008) and the PMN (AraCyc release 8.0; Rhee *et al.*, 2006). Using literature data (Schuster *et al.*, 2006; Knill *et al.*, 2008; Gigolashvili *et al.*, 2009; Li *et al.*, 2011), these reactions were assigned either to the cytosol, to the plastid, or as a transporter into one of those compartments. We also added three reactions (and their associated genes) for benzol-CoA biosynthesis (Kliebenstein *et al.*, 2007; Ibdah *et al.*, 2009; Ibdah & Pichersky, 2009), a compound required for secondary modifications of several methionine-derived glucosinolates. Similarly, while the precise steps of the conversion process between phenylalanine and homophenylalanine are still uncertain in *A. thaliana*, a virtual reaction was added to allow it (Fig. 2a). The reconstructed network contains 43 genes (Table S1) and 144 gene-reaction associations. It includes 121 unique reactions (1 virtual and 91 with at least one gene assigned) and 148 metabolites, including 30 distinct glucosinolates (Table 1).

We merged AraGLS into the core Arabidopsis Metabolic Network (AraGEM; de Oliveira Dal'Molin *et al.*, 2010): the two networks share one reaction (BCAT4) and 29 metabolites. To give proper stoichiometric coefficients, consistent nomenclature, and follow reversibility constraints, it was necessary to add two KEGG reactions to AraGEM and to modify one reaction (see the Description section). Interestingly, the *A. thaliana* gene AT1G73600 (a putative methyltransferase) is coexpressed with the five genes of the core synthesis apparatus for aliphatic glucosinolates (ATTED-II, release 6.0; Mugford *et al.*, 2009). Thus, this gene is listed in Table S1 along with these five genes. It is not, however, included in any network analyses.

Evolutionary history of the glucosinolate pathways

Using comparative genomics, we explored the gene duplications that gave rise to *A. thaliana*'s glucosinolate synthesis pathways. As is illustrated in Fig. 1(a), the first step is catalyzed by the CYP79 family. CYP79A/B is a very old duplication shared with *Carica papaya* (papaya); however these ancestral two enzymes have subsequently reduplicated and diverged functionally in *A. thaliana*. CYP79B2/3 are At-α duplicates while CYP79A2/3 and CYP79F2/3 are post-At-α duplicates. The next synthesis step is

Table 1 The 30 final glucosinolate species produced by the full network

Glucosinolate	Common name	Pathway
3-methylthiopropyl	Glucoiberverin	Homomethionine
3-methylsulfinylpropyl	Glucoiberin	
3-hydroxypropyl		
3-sinapoyloxypropyl		
3-benzoyloxypropyl	Glucomalcommin	
2-propenyl	Sinigrin	
4-methylthiobutyl	Glucoerucin	Dihomomethionine
4-methylsulfinylbutyl	Glucoraphanin	
4-hydroxybutyl		
4-benzoyloxybutyl		
4-sinapoyloxybutyl		
3-butenyl	Gluconapin	
2-hydroxy-3-butenyl	Progoitrin	
2-benzoyloxy-3-butenyl		
2-sinapoyloxy-3-butenyl		
5-methylthiopentyl	Glucoberteroin	Trihomomethionine
5-methylsulfinylpentyl	Glucoalyssin	
4-pentenyl	Glucobrassicanapin	
6-methylthiohexyl	Glucosquerellin	Tetrahomomethionine
6-methylsulfinylhexyl	Glucohesperin	
7-methylthioheptyl		Pentahomomethionine
7-methylsulfinylheptyl	Glucoibarin	
8-methylthiooctyl		Hexahomomethionine
8-methylsulfinyloctyl	Glucohirsutin	
indolylmethyl	Glucobrassicin	Tryptophan
1-methoxy-3-indolylmethyl	Neoglucobrassicin	
4-hydroxy-3-indolylmethyl	4-hydroxygluco-brassicin	
4-methoxy-3-indolylmethyl	4-methoxygluco-brassicin	
benzyl	Glucotropaeolin	Phenylalanine
2-Phenylethyl	Gluconasturtiin	Homophenylalanine

carried out by CYP83B1 and CYP83A1, the products of a pre-α duplication. The glutathione S-transferase step involves four genes: GSTF9 and GSTF10 result from a post-α duplication, while GSTF11 and GSTU20 are independent enzymes performing the same reaction. No duplications are observed in the next two steps, but the second to last step, UDP-glucosyl transferase, is catalyzed by a pair of pre-At-α duplicates (UGT74B1 and C1). Three genes perform the final step: SOT16/SOT17 and SOT18 result from a pre-α tandem duplication, while SOT16 and SOT17 are subsequent At-α duplicates. These duplication patterns are consistent with the metabolomics of glucosinolates in the two species: *C. papaya* possesses only the genes for synthesizing aromatic glucosinolates, including genes orthologous to CYP79 (CP00136G00690 and CP00213G00030), CYP83 (CP00026G02420), C-S lyase (CP00083G00820), UGT74 (CP00005G03360) and ST5 (CP00003G02460). For reasons that are unclear, orthologs of GST and GGP are not known in *C. papaya*. Given the functional distribution of the duplicates in Fig. 1(a), it appears safe to conclude that the fact that *C. papaya* lacks the biosynthetic pathways for indole and methionine-derived aliphatic glucosinolates is partly because of the fact that it does not share the At-β and At-α duplication events (Ming *et al.*, 2008) from which these pathways partly derive. However, one caveat to our argument is that the limited number of At-β syntenic blocks in *A. thaliana* makes it difficult to identify retained duplicates from that whole-genome duplication. Thus, it is possible that some old pre-α duplicates (e.g. the CYP83s) are actually At-β duplicates, particularly because of their absence in *C. papaya*. The community would benefit from a post-At-β and pre-At-α Brassicales reference genome that would permit a more accurate assignment of the duplication histories for the genes in the *A. thaliana* lineage.

Patterns of metabolic flux change under glucosinolate production

We investigated the global changes in flux seen in the AraGEM v1.2 metabolic model when glucosinolate biosynthesis was required (Fig. S1A). Our goal is to optimize the production of new cells (plant biomass), described by a biomass reaction, from a given input of light energy (photons). Here, we employed two biomass reactions for *A. thaliana* Col-0, one previously described (Bekaert *et al.*, 2011) and a second that adds the glucosinolate masses given by Brown *et al.* (2003) to that previous reaction (Table S2). Then, using flux-balance analysis, we searched for flux distributions that maximized biomass production for each equation (see the Description section). The model predicts surprisingly large changes in the flux distribution when glucosinolates are included. In the absence of glucosinolates, 284 (17.3%) of the 1641 internal reactions in the network showed flux, whereas, when all 30 glucosinolates are synthesized, this number increases to 422 (25.7%), including all 121 reactions of the glucosinolate pathways. Moreover, as indicated by Fig. S1(A), production of the 30 distinct glucosinolates also alters the flux for 437 reactions (26.6%) compared with the original model: 196 reactions that are only used when glucosinolate synthesis occurs and 241 that change in flux.

The addition of glucosinolates to the biomass description improves the model's fit to experimental data

The ratio of total nitrogen to total sulfur in *A. thaliana* has been measured to be *c.* 15 : 1 (van der Kooji *et al.*, 1997). However, the biomass reaction from our previous work (Bekaert *et al.*, 2011) gave an uptake ratio of 199 nitrogen atoms to 1 sulfur atom. Adding 13 glucosinolate compounds (Brown *et al.*, 2003) to the leaf biomass reaction reduces this ratio to 10 : 1, a much more reasonable figure, especially as this leaf-only model likely underestimates the true N : S ratio in the whole plant.

Patterns of predicted flux are consistent with gene expression patterns

To evaluate the model under more realistic conditions, we collected the expression status of all network genes (sample GEO: GSM77061; Blasing *et al.*, 2005) in a growing rosette leaf (Brown *et al.*, 2003; Schuster *et al.*, 2006). We disallowed flux through the 51 reactions where the associated genes were not expressed under these conditions (including AOP2 discussed earlier). The resulting metabolic contrasts are illustrated in Fig. S1(B). The optimal

solution in the absence of glucosinolate production predicted a flux in 324 (19.7%) reactions, while in presence of glucosinolates, the model predicted a flux in 412 (25.0%) reactions. A change in flux occurred in 236 nodes not including the glucosinolate pathways.

Cost of glucosinolate synthesis

Our central question is that of the overall energy cost of glucosinolate production during plant growth. As indicated in Table 2, glucosinolate synthesis increases the photon uptake by 17% relative to the base model. In other words, to synthesize the same mass of plant cells with the addition of the glucosinolate masses observed by Brown *et al.* (2003) requires an extra 17 photons per 100 compared with the number in the original system. These costs can best be understood by calculating the quantity of nonglucosinolate biomass produced for a fixed input of photons. We find that the O_2 production is similar in both scenarios, but the importation of sulfur greatly increases when glucosinolates are produced. More importantly, the rate of biomass synthesis decreases by *c.* 15% under those conditions, with a lesser decrease in the import of carbon and nitrogen. Thus, glucosinolate production drains away a meaningful percentage of the energy available to the growing rosette leaves.

The lack of AOP2 and AOP3 in *A. thaliana* Columbia (Col-0) appears to result in lowered aliphatic glucosinolate production (Kliebenstein *et al.*, 2001). Similarly, the concentrations of glucosinolates reported for this strain by Brown *et al.* (2003) were taken in the absence of glucosinolate induction by predation. Thus, although our model is intended to describe this strain, it is reasonable to ask if the estimated costs respond predictably to the increased glucosinolate production characteristic of other lines or resulting from induction in this line. We thus doubled the concentration of each of the 13 glucosinolates in the model's biomass description. As would be expected from the linear models underlying FBA, the result was a doubling of the photon cost of glucosinolate production (data not shown).

Table 2 Import and export flux of the main metabolites for a fixed amount of biomass production (top) and fixed number of input photons (bottom)

Flux	Leaf*	Leaf + GLS*	Variation (fold)
Biomass	1000.0	1000.0	1.00
Exported oxygen	22 013	25 317.9	1.15
Imported CO_2	23 734.6	26 943.7	1.14
Imported H_2O	45 526.7	59 201.5	1.30
Imported hydrogen sulfide	24.9	355.3	14.3
Imported NH_3	4964.32	5529.7	1.11
Imported sulfate	0.0	179.2	∞
Imported photons	177 000	207 765	1.17
Biomass	1000.0	851.924	0.85
Export oxygen	22 013	21 568.9	0.98
Import CO_2	23 734.6	22 954	0.97
Import H_2O	45 526.7	50 435.2	1.11
Import hydrogen sulfide	24.9	302.689	12.16
Import NH_3	4964.32	4710.88	0.95
Import sulfate	0.0	152.665	∞
Import photons	177 000	177 000	1.00

*Arbitrary units.

The suite of glucosinolates produced by *A. thaliana* is relatively inexpensive

Arabidopsis thaliana produces a complex set of glucosinolates. We were curious if this particular combination of glucosinolates was especially energetically favorable or unfavorable. We thus randomly reassigned the values of glucosinolates requirements between the 13 glucosinolates present in the biomass reaction and recalculating the flux and the photon uptake for 1000 permutations (Fig. 3). Effectively, the molar quantities of total glucosinolates were thus unchanged. The set of glucosinolates produced by the plant is relatively less costly (in terms of photons consumed) than are most equivalent random sets ($P = 0.12$).

Discussion

Glucosinolates are a model for understanding the evolution of herbivore resistance (Rausher, 2001). Two key questions about such defenses have to do with the degree to which the defense also imposes costs on the host and whether or not the host and its herbivores coevolve. With regard to the second question, there is now evidence for such coevolution, coming from the biochemistry of the herbivores (Wheat *et al.*, 2007) and from the existence of natural variation in these traits that is coupled to geography (Bidart-Bouzat & Kliebenstein, 2008; Todesco *et al.*, 2010). Ironically, increased glucosinolate production can drive increased predation by specialized insects adapted to these compounds (Bidart-Bouzat & Kliebenstein, 2008), suggesting that the coevolution may be ongoing. At a minimum, such selection must have been active in the relatively recent past, given the evolutionary radiation of glucosinolate synthesis enzymes after the *Arabidopsis* WGB-α event (Fig. 1).

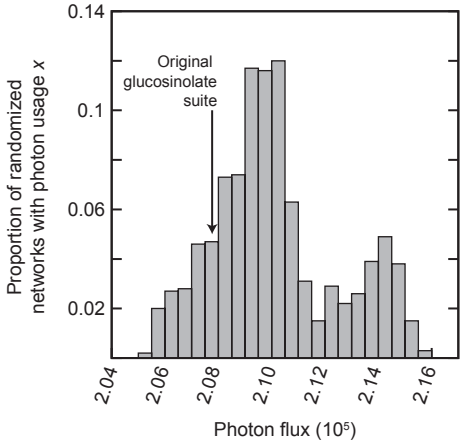

Fig. 3 Distribution of the photon (arbitrary units) uptake for 1000 permutations of the concentrations of the glucosinolates in the biomass reaction. The *x*-axis gives the relative number of photons needed to synthesize 1000 (arbitrary) units of biomass. The photon usage for the original network (207 765) is Indicated by an arrow. Only 123 of 1000 randomized glucosinolate suites required as few or fewer photons to synthesize 1000 units of biomass.

This study aimed to synthesize and integrate glucosinolate metabolism into the core *A. thaliana* metabolic network and to investigate the energetic costs associated with their production. Significant progress has been made in recent years towards understanding the biochemistry and molecular genetics of the biosynthesis of glucosinolates (Textor & Gershenzon, 2009; Sønderby *et al.*, 2010) and most, if not all, of the enzymes are now known (43 genes; Table S1). However, details for several of the reactions are still incomplete, with the stoichiometry of certain reactions in particular being somewhat uncertain (e.g. AOP2 production or the conversion mechanism from phenylalanine to homophenylalanine). Their integration into the core network (Fig. 2) is the first step in understanding both the plant's metabolism and its defense against insects. Currently, the glucosinolate pathways are directly linked to the primary network by three amino acids (methionine, phenylalanine and tryptophan) and glutathione, but this isolation is partly because secondary metabolism is incompletely incorporated into the network, making it difficult to fully appreciate how these pathways interact with the rest of the cell.

We found that glucosinolate production at the concentration observed in *A. thaliana* Col-0 has a large and unavoidable energy allocation cost, roughly equivalent to 15% of the total energy needed to synthesize the other components of a leaf cell (Table 2). (For reference, Table 3 gives the synthesis costs of the three amino acids involved in glucosinolate synthesis.) Moreover, this estimate is likely something of a lower bound, given that other lineages of *A. thaliana* synthesize more glucosinolates and such synthesis can be induced by predation (Kliebenstein *et al.*, 2001; Textor & Gershenzon, 2009). Our results are also in keeping with recent experimental work in *Brassica rapa* (field mustard) showing that the cost of glucosinolate synthesis is manifest in reduced numbers of seeds and flowers per plant at higher levels of glucosinolate synthesis (Stowe & Marquis, 2011), traits that could well be linked to amounts of metabolic 'investment'. In plants, fitness is more closely related to the number of successful offspring produced than to the production of new biomass, and it is not clear to what extent energy costs limit plant reproduction. Our argument is thus not that we have estimated a fitness cost, but rather that we find it difficult to accept that energy is so plentiful for the plant that an expenditure of this magnitude is irrelevant. (Note, however, that the current metabolic model does not incorporate the energy required for homeostasis, suggesting that the relative cost of glucosinolates is lower than described here.) Our findings are consistent with a growth rate analysis of knockout mutants for various glucosinolate biosynthesis genes that revealed a negative correlation between glucosinolate concentrations and growth rates during early developmental stages (Orth *et al.*, 2010; Züst *et al.*, 2011). These data collectively support the idea that glucosinolate biosynthesis is costly. The knockout mutants had a significant competitive advantage in the absence of herbivores, an advantage that even allowed them to flower earlier than wild-type plants.

Table 3 Cost of the drained amino acids (for one unit of each amino acid)

Flux	L-Tryptophan ($C_{11}H_{12}N_2O_2$)	L-Phenylalanine ($C_9H_{11}NO_2$)	L-Methionine ($C_5H_{11}NO_2S$)
Export oxygen	11.5	10	5.5
Import CO_2	11	9	5
Import H_2O	37.5	34	19.5
Import hydrogen sulfide	0	0	1
Import NH_3	2	1	1
Import photons	92	80	44

It is also important to recall that our analysis is based on idealized conditions for the plant where all nutrients, including the sulfur and nitrogen building blocks of the glucosinolates, are provided in excess. Instead, the modeled growth rates are limited by the supply of light energy. Similarly, the model does not actually distinguish individual cells in the plant: the fluxes account instead for the movement of raw materials (mass) into the plant and their use in biosynthesis. The result of these assumptions is to make potential mass limitations energy limitations: a real plant cell would likely find glucosinolate production even more expensive depending on the nutrient environment it finds itself in, having to (at best) spend even more energy to import the necessary building blocks. There is considerable evidence that such elemental limitations are significant for real plants: to cite just one example, sulfur assimilation genes are coregulated with the glucosinolate pathways in *A. thaliana* (Yatusevich *et al.*, 2010). Note also that, although it is natural to think of the cost associated with defense as being metabolic, defense is also costly because the induced compounds may be damaging to the host (Strauss *et al.*, 2002; Todesco *et al.*, 2010) as well as having other indirect costs acting on the plant's physiology or ecology (Stowe, 1998; Purrington & Bergelson, 1999; Strauss *et al.*, 2002). As a result, unlike microbes where metabolic predictions can be directly translated into fitness costs that can be experimentally tested (at least in a strain competition scenario; Dekel & Alon, 2005), it is difficult to develop experiments that more directly link metabolism to fitness in plants and, indeed, we would certainly not make such a claim for this work.

On the other hand, further evidence for the costly nature of glucosinolate production comes from a surprising source: although synthesis is costly, the particular suite of glucosinolates observed is actually relatively inexpensive relative to most other potential combinations (Fig. 3). Given these costs and patterns, it is also difficult to conceive of a nonadaptive role that has maintained this production over Brassicaceae evolution. The complexity of the glucosinolate suite in *A. thaliana* also recalls the fact that certain glucosinolates can actually attract specialized herbivores (Bidart-Bouzat & Kliebenstein, 2008), and we suggest that the continued evolution of the herbivores has driven the plant to continuously adjust its suite of glucosinolates. The pathways to glucosinolate diversity seem to have evolved quite recently: many initial duplications were created by the late Cretaceous At-α event (46.9–64.5 mya; Beilstein *et al.*, 2010), with subsequent duplications helping to specialize these genes (Fig. 1). Notably, the glucosinolate-feeding Pierina butterflies appear to have undergone a contemporaneous radiation, further suggesting a pattern of coevolution between the Pierid herbivores and their Brassicaceae hosts (Wheat *et al.*, 2007; Beilstein *et al.*, 2010; Schranz *et al.*, 2011).

Systems biology approaches to evolution are attractive as they bring the promise of quantitative analyses of trait evolution, integrating phenomena from the molecular to the ecosystem level. As a result, they allow us to describe precise hypotheses as to how traits are adaptive. Although metabolic models of plants cannot yet directly predict fitness, the ability to link predictive models to systems that have been carefully studied in the laboratory and the field will prove powerful. It is particularly noteworthy that a complex and evolutionarily costly system such as that of glucosinolates has evolved in a relatively short period of time, using some of the raw material already in the genome (e.g. Fig. 1). There now remains little doubt that the driver of this evolution was selection in favor of adoption of protective compounds, even if those compounds required significant metabolic investment. This fact reminds us again that fitness is not an absolute concept but a particular set of tradeoffs appropriate to the organism's local environments.

Acknowledgements

We would like to thank T. Arias, D. Mayfield, A. Ferrieri, C. Coffman, and J. Schultz for helpful discussions. This work was supported by a Research Board grant from the University of Missouri (M.B.), a National Library of Medicine Biomedical and Health Informatics Training Fellowship (LM007089-19) (C.M. H.), the Reproductive Biology Group of the Food for the 21st Century program at the University of Missouri (G.C.C.) and by the US National Science Foundation (DBI 0501712 and DBI 0638536) (P.E. and J.C.P.).

References

Agrawal AA. 2011. New synthesis-trade-offs in chemical ecology. *Journal of Chemical Ecology* **37**: 230–231.

Agrawal AA, Fishbein M. 2006. Plant defense syndromes. *Ecology* **87**: S132–S149.

Agrawal AA, Heil M. 2012. Synthesizing specificity: multiple approaches to understanding the attack and defense of plants. *Trends in Plant Science* **17**: 239–242.

Bastian M, Heymann S, Jacomy M 2009. Gephi: an open source software for exploring and manipulating networks. In: *Third International AAAI Conference on Weblogs and Social Media*. San Jose, CA, USA: AAAI Publications, 361–362.

Beekwilder J, van Leeuwen W, van Dam NM, Bertossi M, Grandi V, Mizzi L, Soloviev M, Szabados L, Molthoff JW, Schipper B, *et al.* 2008. The impact of the absence of aliphatic glucosinolates on insect herbivory in *Arabidopsis*. *PLoS ONE* **3**: e2068.

Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S. 2010. Dated molecular phylogenies indicate a miocene origin for *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* **107**: 18724–18728.

Bekaert M, Edger PP, Pires JC, Conant GC. 2011. Two-phase resolution of polyploidy in the *Arabidopsis* metabolic network gives rise to relative and absolute dosage constraints. *The Plant Cell* **23**: 1719–1728.

Bidart-Bouzat MG, Kliebenstein DJ. 2008. Differential levels of insect herbivory in the field associated with genotypic variation in glucosinolates in *Arabidopsis thaliana*. *Journal of Chemical Ecology* **34**: 1026–1037.

Blanc G, Hokamp K, Wolfe KH. 2003. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. *Genome Research* **13**: 137–144.

Blasing OE, Gibon Y, Gunther M, Hohne M, Morcuende R, Osuna D, Thimm O, Usadel B, Scheible WR, Stitt M. 2005. Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. *The Plant Cell* **17**: 3257–3281.

Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**: 433–438.

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J. 2003. Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. *Phytochemistry* **62**: 471–481.

Burow M, Markert J, Gershenzon J, Wittstock U. 2006. Comparative biochemical characterization of nitrile-forming proteins from plants and insects that alter myrosinase-catalysed hydrolysis of glucosinolates. *The FEBS Journal* **273**: 2432–2446.

Catoni E, Schwab R, Hilpert M, Desimone M, Schwacke R, Flugge UI, Schumacher K, Frommer WB. 2003. Identification of an *Arabidopsis* mitochondrial succinate-fumarate translocator. *FEBS Letters* **534**: 87–92.

Csardi G, Nepusz T. 2006. The igraph software package for complex network research. *InterJournal* Complex Systems: 1695.

Dekel E, Alon U. 2005. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**: 588–592.

Edwards JS, Palsson BO. 2000. The *Escherichia coli* mg1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences, USA* **97**: 5528–5533.

Elena SF, Lenski RE. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics* **4**: 457–469.

Fahey JW, Zalcmann AT, Talalay P. 2001. The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. *Phytochemistry* **56**: 5–51.

Gigolashvili T, Engqvist M, Yatusevich R, Muller C, Flugge UI. 2008. HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. *New Phytologist* **177**: 627–642.

Gigolashvili T, Yatusevich R, Berger B, Muller C, Flugge UI. 2007. The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. *The Plant Journal* **51**: 247–261.

Gigolashvili T, Yatusevich R, Rollwitz I, Humphry M, Gershenzon J, Flugge UI. 2009. The plastidic bile acid transporter 5 is required for the biosynthesis of methionine-derived glucosinolates in *Arabidopsis thaliana*. *The Plant Cell* **21**: 1813–1829.

Halkier BA, Gershenzon J. 2006. Biology and biochemistry of glucosinolates. *Annual Review of Plant Biology* **57**: 303–333.

Hall JC, Iltis HH, Sytsma KJ. 2004. Molecular phylogenetics of core brassicales, placement of orphan genera *emblingia*, *forchhammeria*, *tirania*, and character evolution. *Systematic Botany* **29**: 654–669.

Hartl DL, Clark AG. 1997. *Principles of population genetics*. Sunderland, MA, USA: Sinauer Associates.

Ibdah M, Chen YT, Wilkerson CG, Pichersky E. 2009. An aldehyde oxidase in developing seeds of *Arabidopsis* converts benzaldehyde to benzoic acid. *Plant Physiology* **150**: 416–423.

Ibdah M, Pichersky E. 2009. *Arabidopsis* Chy1 null mutants are deficient in benzoic acid-containing glucosinolates in the seeds. *Plant Biology* **11**: 574–581.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, *et al.* 2008. Kegg for linking genomes to life and the environment. *Nucleic Acids Research* **36**: D480–D484.

Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K. 2007. The atgenexpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *The Plant Journal* **50**: 347–363.

Klein M, Reichelt M, Gershenzon J, Papenbrock J. 2006. The three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed. *The FEBS Journal* **273**: 122–136.

Kliebenstein DJ, D'Auria JC, Behere AS, Kim JH, Gunderson KL, Breen JN, Lee G, Gershenzon J, Last RL, Jander G. 2007. Characterization of seed-specific benzoyloxyglucosinolate mutations in *Arabidopsis thaliana*. *The Plant Journal* **51**: 1062–1076.

Kliebenstein DJ, Kroymann J, Mitchell-Olds T. 2005. The glucosinolate-myrosinase system in an ecological and evolutionary context. *Current Opinion in Plant Biology* **8**: 264–271.

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T. 2001. Gene duplication in the diversification of secondary metabolism: tandem

2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in *Arabidopsis*. *The Plant Cell* **13**: 681–693.

Knill T, Schuster J, Reichelt M, Gershenzon J, Binder S. 2008. *Arabidopsis* branched-chain aminotransferase 3 functions in both amino acid and glucosinolate biosynthesis. *Plant Physiology* **146**: 1028–1039.

van der Kooji TAW, de Kok LJ, Haneklaus S, Schnug E. 1997. Uptake and metabolism of sulphur dioxide by *Arabidposis thaliana*. *New Phytologist* **135**: 101–107.

Li J, Kristiansen KA, Hansen BG, Halkier BA. 2011. Cellular and subcellular localization of flavin-monooxygenases involved in glucosinolate biosynthesis. *Journal of Experimental Botany* **62**: 1337–1346.

Loewe L. 2009. A framework for evolutionary systems biology. *BMC System Biology* **3**: 27.

Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences, USA* **104**: 8597–8604.

Lyons E, Freeling M. 2008. How to usefully compare homologous plant genes and chromosomes as DNA sequences. *The Plant Journal* **53**: 661–673.

Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, Wang X, Bowers J, Paterson A, Lisch D, *et al.* 2008. Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: coge with rosids. *Plant Physiology* **148**: 1772–1781.

Mauricio R. 1998. Costs of resistance to natural enemies in field populations of the annual plant *Arabidopsis thaliana*. *The American Naturalist* **151**: 20–28.

Mauricio R, Rausher MD. 1997. Experimental manipulation of putative selective agents provides evidence for the role of natural enemies in the evolution of plant defense. *Evolution* **51**: 1435–1444.

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, *et al.* 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* **452**: 991–996.

Mithen R, Bennett R, Marquez J. 2010. Glucosinolate biochemical diversity and innovation in the brassicales. *Phytochemistry* **71**: 2074–2086.

Mugford SG, Yoshimoto N, Reichelt M, Wirtz M, Hill L, Mugford ST, Nakazato Y, Noji M, Takahashi H, Kramell R, *et al.* 2009. Disruption of adenosine-5'-phosphosulfate kinase in *Arabidopsis* reduces levels of sulfated secondary metabolites. *The Plant Cell* **21**: 910–927.

de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK. 2010. Aragem, a genome-scale reconstruction of the primary metabolic network in *Arabidopsis*. *Plant Physiology* **152**: 579–589.

Orth JD, Thiele I, Palsson BO. 2010. What is flux balance analysis? *Nature Biotechnology* **28**: 245–248.

Pons P, Latapy M 2005. Computing communities in large networks using random walks. In: Yolum P, Gungor T, Gurgen F, Ozturan C, eds. *Computer and Information Sciences – ISCIS 2005*. Berlin, Heidelberg, Germany: Springer, 284–293.

Purrington CB, Bergelson J. 1999. Exploring the physiological basis of costs of herbicide resistance in *Arabidopsis thaliana*. *The American Naturalist* **154**: 82–91.

Rausher MD. 2001. Co-evolution and plant resistance to natural enemies. *Nature* **411**: 857–864.

Rhee SY, Zhang P, Foerster H, Tissier C. 2006. Aracyc: overview of an *Arabidopsis* metabolism database and its applications for plant research. *Plant Metabolomics* **57**: 141–154.

Rodman JE, Soltis PS, Soltis DE, Sytsma KJ, Karol KG. 1998. Parallel evolution of glucosinolate biosynthesis inferred from congruent nuclear and plastid gene phylogenies. *American Journal of Botany* **85**: 997–1006.

Ronse De Craene LP, Haston E. 2006. The systematic relationships of glucosinolate-producing plants and related families: a cladistic investigation based on morphological and molecular characters. *Botanical Journal of the Linnean Society* **151**: 453–494.

Schranz ME, Edger PP, Pires JC, van Dam NM, Wheat CW 2011. Comparative genomics in the brassicales: ancient genome duplications, glucossinolate diversification and pierinae herbivore radiation. In: Edwards JBD, Parkin I, Kole C, eds. *Genetics, genomics and breeding of oilseed brassicas*. Jersey, UK: Science Publishers, 206–218.

Schuster J, Knill T, Reichelt M, Gershenzon J, Binder S. 2006. Branched-chain aminotransferase4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in *Arabidopsis*. *The Plant Cell* **18**: 2664–2679.

Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E. 2008. Network-based prediction of human tissue-specific metabolism. *Nature Biotechnology* **26**: 1003–1010.

Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Van de Peer Y. 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA* **99**: 13627–13632.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* **96**: 336–348.

Sønderby IE, Geu-Flores F, Halkier BA. 2010. Biosynthesis of glucosinolates – gene discovery and beyond. *Trends in Plant Science* **15**: 283–290.

Sønderby IE, Hansen BG, Bjarnholt N, Ticconi C, Halkier BA, Kliebenstein DJ. 2007. A systems biology approach identifies a R2R3 MYB gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. *PLoS ONE* **2**: e1322.

Stoebel DM, Dean AM, Dykhuizen DE. 2008. The cost of expression of *Escherichia coli* lac operon proteins is in the process, not in the products. *Genetics* **178**: 1653–1660.

Stowe KA. 1998. Experimental evolution of resistance in *Brassica rapa*: correlated response of tolerance in lines selected for glucosinolate content. *Evolution* **52**: 703–712.

Stowe KA, Marquis RJ. 2011. Costs of defense: correlated responses to divergent selection for foliar glucosinolate content in *Brassica rapa*. *Evolutionary Ecology* **25**: 763–775.

Strauss SY, Rudgers JA, Lau JA, Irwin RE. 2002. Direct and ecological costs of resistance to herbivory. *Trends in Ecology & Evolution* **17**: 278–285.

Textor S, Gershenzon J. 2009. Herbivore induction of the glucosinolate–myrosinase defense system: major trends, biochemical bases and ecological significance. *Phytochemistry Reviews* **8**: 149–170.

Thomas BC, Pedersen B, Freeling M. 2006. Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome research* **16**: 934–946.

Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M, Epple P, Kuhns C, Sureshkumar S, Schwartz C, Lanz C, *et al.* 2010. Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* **465**: 632–636.

Wagner A. 2005. Energy constraints on the evolution of gene expression. *Molecular Biology and Evolution* **22**: 1365–1374.

Wheat CW, Vogel H, Wittstock U, Braby MF, Underwood D, Mitchell-Olds T. 2007. The genetic basis of a plant-insect coevolutionary key innovation. *Proceedings of the National Academy of Sciences, USA* **104**: 20427–20431.

Windsor AJ, Reichelt M, Figuth A, Svatos A, Kroymann J, Kliebenstein DJ, Gershenzon J, Mitchell-Olds T. 2005. Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (brassicaceae). *Phytochemistry* **66**: 1321–1333.

Wright J, Wagner A. 2008. The systems biology research tool: evolvable open-source software. *BMC systems biology* **2**: 55.

Yatusevich R, Mugford SG, Matthewman C, Gigolashvili T, Frerigmann H, Delaney S, Koprivova A, Flugge UI, Kopriva S. 2010. Genes of primary sulfate assimilation are part of the glucosinolate biosynthetic network in *Arabidopsis thaliana*. *The Plant Journal* **62**: 1–11.

Züst T, Joseph B, Shimizu KK, Kliebenstein DJ, Turnbull LA. 2011. Using knockout mutants to reveal the growth costs of defensive traits. *Proceedings of the Royal Society B: Biological Sciences* **278**: 2598–2603.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Flux induced by the implementation of glucosinolate pathways.

Table S1 Gene inventory of the glucosinolate pathway

Table S2 Biomass reaction

Notes S1 AraGEM v1.2 metabolic network in System Biology Markup Language (SBML) format.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.