
This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of May 1, 2014):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/344/6183/510.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2014/04/30/344.6183.510.DC1.html>

This article **cites 48 articles**, 21 of which can be accessed free:

<http://www.sciencemag.org/content/344/6183/510.full.html#ref-list-1>

This article appears in the following **subject collections**:

Botany

<http://www.sciencemag.org/cgi/collection/botany>

Evolution

<http://www.sciencemag.org/cgi/collection/evolution>

Genomic Signatures of Specialized Metabolism in Plants

Lee Chae, Taehyong Kim, Ricardo Nilo-Poyanco, Seung Y. Rhee*

All plants synthesize basic metabolites needed for survival (primary metabolism), but different taxa produce distinct metabolites that are specialized for specific environmental interactions (specialized metabolism). Because evolutionary pressures on primary and specialized metabolism differ, we investigated differences in the emergence and maintenance of these processes across 16 species encompassing major plant lineages from algae to angiosperms. We found that, relative to their primary metabolic counterparts, genes coding for specialized metabolic functions have proliferated to a much greater degree and by different mechanisms and display lineage-specific patterns of physical clustering within the genome and coexpression. These properties illustrate the differential evolution of specialized metabolism in plants, and collectively they provide unique signatures for the potential discovery of novel specialized metabolic processes.

Plants produce compounds that contribute to human health and survival. More than one-third of human drugs have originated from products synthesized by plant metabolism (1). Understanding the origins, evolution, and diversity of plant metabolites has been a long-standing goal in plant biology. Metabolic diversification may play a key role in plant evolution, particularly those genes that encode proteins that function as novel specialized (or secondary) metabolites that mediate interactions between a given plant and its environment (2, 3).

In order to apply advances in genomic and computational resources, both systematically and quantitatively, to assess the relationship between metabolic diversity and plant evolution (4, 5), we performed a comparative analysis of the metabolic

networks of 16 green plants with high-quality genome annotations from major groups in the green plant lineage. These included six chlorophyte algae (*Chlamydomonas reinhardtii*, *Volvox carteri*, *Coccomyxa subellipsoidea*, *Micromonas pusilla* CCMP1545 and RCC299, *Ostreococcus lucimarinus*), two early-diverging land plants [*Physcomitrella patens* (moss), *Selaginella moellendorffii*], three grasses [*Oryza sativa* ssp. *japonica* (rice), *Sorghum bicolor*, *Zea mays* (maize)], and five eudicots [*Arabidopsis thaliana*, *Manihot esculenta* (cassava), *Populus trichocarpa* (poplar), *Glycine max* (soybean), *Vitis vinifera* (grapevine)] (6) (table S1 and fig. S1). We annotated enzymes encoded in the genomes of the 16 species with four-part Enzyme Commission (EC) numbers (6), implementing an ensemble classification algorithm to integrate molecular function predictions (fig. S2) (6). Pipeline validation produced results of 86% precision, 87% recall, and 86% F_1 measure, compared to the individual methods (54 to 81%

precision, 57 to 82% recall, and 55 to 79% F_1 measure) (figs. S3 and S4). The identified enzyme inventories ranged in size from 1295 in *Ostreococcus* (16.6% of protein-coding genes) to 12,124 in soybean (21.7% of protein-coding genes), covering 1219 distinct EC numbers (table S2, fig. S5, and data file S1). Annotated enzymes were used to reconstruct bidirectional, reaction-centric metabolic networks (7). Enzymes and reactions were manually classified into 13 functional classes, including primary metabolism-related categories such as amino acid, nucleotide, and energy metabolism, as well as specialized metabolism (fig. S6 and data file S2) (6).

To assess divergence among the genome-scale metabolic networks, we investigated differences in their structure and content. The size (diameter) and density (clustering coefficient and betweenness centrality distributions) (7) of the metabolic networks were statistically similar across the species (fig. S7). However, land plant networks contain a significantly larger number of reaction nodes than algal networks (fig. S8A). Given the similarity in network size and density, the increase in reaction nodes suggests a more complex set of networks in land plants, with higher levels of connectivity, as seen by differences in the degree and average neighbor degree distributions ($P < 0.0001$, Kruskal-Wallis test) (fig. S8, B and C) (8).

Comparisons of network content, reflecting species phylogeny, grouped the prasinophyceae apart from chlorophyceae algae and separated moss and *Selaginella* from grasses and eudicots, respectively (Fig. 1A). This suggests that the evolution of plant metabolic networks followed a pattern of descent with modification, with closely related species sharing more similar sets of metabolic reactions. If true, phylogenetic comparisons of genome-scale metabolic networks may offer a predictive tool for discovering metabolic properties in different plant

Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA.

*Corresponding author. E-mail: srhee@carnegiescience.edu

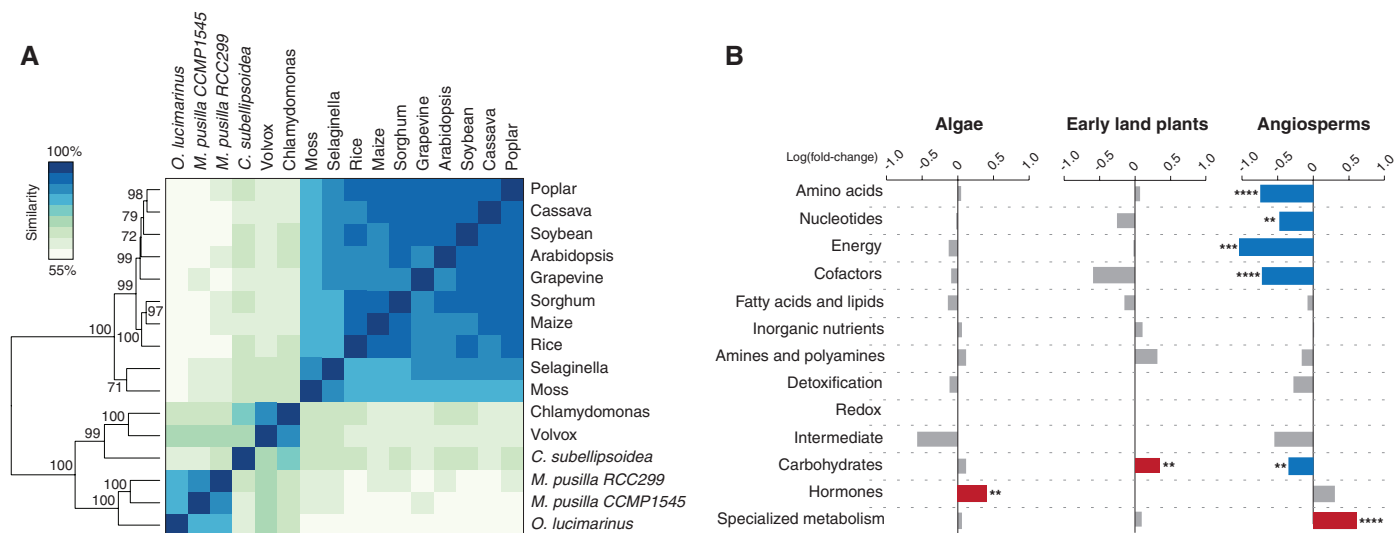


Fig. 1. Diversification of metabolism in the green plant lineage. (A) Heat map depicting the similarity of reaction node sets among species (6). Numbers indicate support from 1000 bootstrap rounds. (B) Enrichment (red) or depletion (blue) of functional classes in lineage-specific groupings of metabolic reactions

(algae, 6 species, 154 reactions; early land plants, 2 species, 32 reactions; angiosperms, 8 species, 137 reactions). The x axes depict \log_{10} transformations of relative change ratios for reactions unique to each group versus all reactions in each group. ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ (hypergeometric test).

species. However, reconstructing events leading to differences in metabolic reaction sets across a deep phylogeny, as represented by the 16 species, is challenged by the paucity of high-quality genome sequence data across the plant kingdom (9).

If metabolic networks evolved by descent with modification, lineage-based groupings of enzymatic reactions may reveal how metabolic functionality diverged after important junctures in plant evolution. We investigated the functional composition of the set of reactions unique to algae, the early land plants, and angiosperms (323 reactions, 19.9% of the network reactions). Algae were enriched in hormone-related reactions ($P = 0.001$, hypergeometric test), whereas early land plant reactions were enriched in carbohydrate metabolism ($P = 0.003$, hypergeometric test) (Fig. 1B). Reactions unique to the angiosperms were enriched for specialized metabolism ($P \approx 0$, hypergeometric test), whereas primary metabolism-related reactions were significantly underrepresented

in this set ($P < 0.01$ for amino acid, carbohydrate, cofactor, energy, and nucleotide metabolism; hypergeometric test).

Several serotonin- and melatonin-related reactions were found in the algal groups examined (data file S3). However, other reactions typically comprising these metazoan pathways, such as the decarboxylation of 5-hydroxytryptophan or the acetylation of serotonin, were not found. Similarly, mosses and bryophytes contained a number of carbohydrate-related reactions not associated with any known pathway (6). The lack of genome sequences for other members of these lineages makes it hard to reconstruct events that may have led to the appearance of these reactions in these species. On one hand, they may reflect a mosaic pattern of metabolic evolution, where ancestral pathways are unevenly conserved across reactions and lineages. Conversely, some proportion of the reactions may represent the appearance of newfound metabolic functionality in a particular species.

The enrichment pattern in the angiosperm grouping suggests that the predominant metabolic innovations after the appearance of the vascular plants dealt with specialized metabolism, while those processes related to primary metabolism were likely established earlier in the evolution of plants. The extent of specialized metabolism evolution in basal plant lineages is difficult to ascertain, as much of our knowledge of plant specialized metabolism derives from angiosperms. Consequently, it is likely that a number of specialized processes unique to basal lineages have not yet been discovered, which ultimately may demonstrate a greater role for specialized metabolism in the basal plant lineages (10).

The functional enrichment analysis suggests that specialized metabolic capacity expanded after the divergence of early land plants into more derived lineages, including the angiosperms. We thus examined whether specialized metabolism exhibited different patterns of evolution among

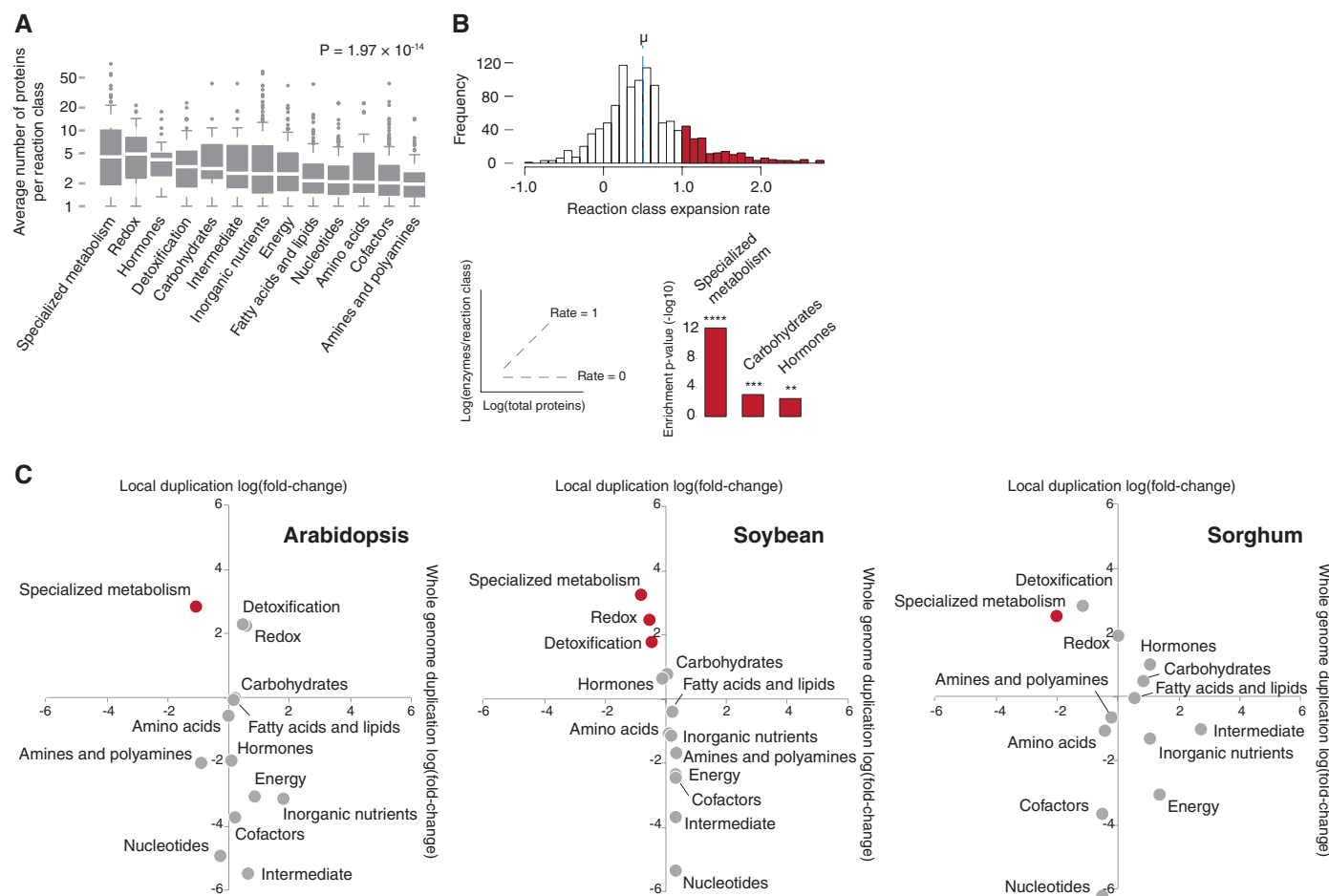


Fig. 2. Differential evolution of specialized metabolism in plants. (A) Differences in the distributions of the number of enzymes per EC-based reaction class for each functional category in all 16 species ($P = 1.97 \times 10^{-14}$, Kruskal-Wallis test). (B) Histogram of expansion rates for all reaction classes found in two or more species (6). Inset is an example plot for one reaction class. Rate = 0, enzyme number remained constant; rate = 1, enzyme number increased linearly with protein number. Specialized metabolism ($P = 5.64 \times 10^{-13}$, hypergeometric test), carbohydrate ($P = 8.40 \times 10^{-4}$, hypergeometric

test), and hormone ($P = 0.003$, hypergeometric test) reactions are enriched in the set of reactions with expansion rates ≥ 1 . The y axis represents negative log₁₀ transformations of the P value. (C) Enrichment or depletion of genes annotated to a given functional class after local duplication (LD, x axis) versus whole-genome duplication (WGD, y axis). Axes represent log₁₀ transformations of relative change ratios for LD or WGD genes in each functional grouping versus all metabolic LD or WGD genes. Red markers indicate LD enrichment and WGD depletion at $P \leq 0.0001$ (hypergeometric test).

enzyme families. Overall, the total number of enzymes observed for each species increased linearly with respect to the total number of proteins ($R^2 = 0.947$, $P = 6.17 \times 10^{-7}$, Student's t test) (fig. S5). However, this increase is biased toward specialized metabolism, as specialized metabolism reactions catalyzed by specific four-digit EC numbers tended to have larger enzyme inventories relative to other functional classes ($P = 1.97 \times 10^{-14}$, Kruskal-Wallis test) (Fig. 2A and table S3). Furthermore, a number of metabolic reactions experienced enzyme number expansions at a rate greater than the rate of change in the total number of proteins (Fig. 2B) (6). Several functional classes were enriched in this set of metabolic reactions, most notably specialized metabolism ($P = 5.64 \times 10^{-13}$, hypergeometric test), along with carbohydrates ($P = 8.40 \times 10^{-4}$, hypergeometric test) and hormones ($P = 0.003$, hypergeometric test). Similar expansions in enzyme number related to specialized metabolism have been observed, notably within some sub-families of the cytochrome P450s and glycosyltransferases (11, 12). These results suggest that specialized metabolism is under selection driving the expansion of gene families coding for specific enzymatic processes in plants, in contrast to genes underlying the function of primary metabolism-related processes.

This growth in enzyme families is the result of selection for gene duplication for specialized metabolic genes. In general, gene duplication events, such as whole-genome duplications (WGDs) and local (tandem) duplication (LDs), increase gene content in plant genomes (13). We investigated the impact of WGD and LD events on amplifying metabolic genes across functional categories. In

Arabidopsis, specialized metabolic genes were the only metabolic functional class significantly enriched in LD genes ($P \approx 0$, hypergeometric test) that were also significantly depleted in the WGD-derived gene set ($P = 4.80 \times 10^{-7}$, hypergeometric test) (Fig. 2C). Similar enrichment of LD genes was reported for three *Arabidopsis* specialized metabolic pathways but suggested that whole-genome duplications contributed significantly to these pathways (14). We tested whether whole-genome duplications have contributed significantly to the specialized metabolic gene sets in soybean and sorghum (6). Both species exhibited significant depletions of WGD-derived specialized metabolic genes ($P = 2.23 \times 10^{-37}$ and 6.77×10^{-14} , respectively; hypergeometric test) but displayed significant enrichment in LD-derived specialized metabolic genes ($P \approx 0$ for both species, hypergeometric test) (Fig. 2C).

Although the LD versus WGD dynamic was the same among the three species, the functional impact of local duplication on specialized metabolic genes differed. For example, a local duplication of genes involved in sinapate ester production was seen in *Arabidopsis*, whereas local duplications of terpenoid genes associated with phytoalexins were observed in sorghum (data files S1 and S4). These duplications appeared to affect a number of flavonol-related genes in each species, including those associated with quercetin and rhusin in *Arabidopsis* and kaempferol in soybean (data files S1 and S4). Thus, the expansion of specialized metabolic plant genes appears to be due to local gene duplication.

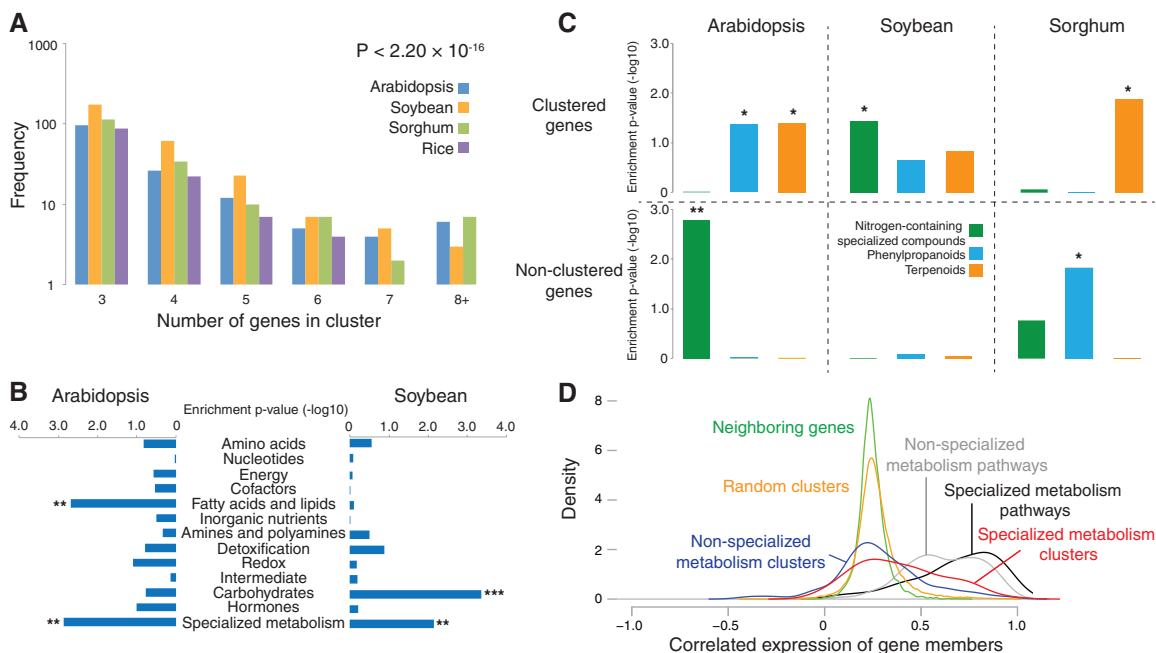
We also examined the genomic colocation, or clustering, of genes whose products function together to synthesize specialized metabolites (15). Reports thus far have focused on specific meta-

bolic gene clusters, but it is unknown whether clustering plays a major role in specialized metabolism evolution. Consequently, we analyzed the extent of metabolic gene clustering and tested whether clusters are biased for specialized metabolism (6). From this we observed that approximately one-third of the metabolic genes in *Arabidopsis* (30.1%), soybean (30.2%), and sorghum (30.5%), and one-fifth of the genes in rice (22.4%), were situated in clusters (data file S5). For each species, the extent of clustering was greater than expected by chance ($P < 2.20 \times 10^{-16}$, χ^2 test) (Fig. 3A) (6). The *Arabidopsis* and soybean clusters were significantly enriched for specialized metabolic genes ($P = 1.38 \times 10^{-3}$ and 7.06×10^{-3} , respectively; hypergeometric test), as well as fatty acids and lipids in *Arabidopsis* ($P = 2.01 \times 10^{-3}$, hypergeometric test) and carbohydrates in soybean ($P = 4.34 \times 10^{-4}$, hypergeometric test) (Fig. 3B). In contrast, no such enrichment was found in sorghum or rice (table S4).

Clustering patterns differed across species and different classes of specialized metabolic compounds (6). Genes involved in phenylpropanoid and terpenoid metabolism were more likely to be found in *Arabidopsis* clusters ($P = 0.043$ and 0.040 , hypergeometric test), whereas genes involved in nitrogen-containing specialized compound metabolism were enriched in the nonclustered gene set ($P = 0.002$, hypergeometric test) (Fig. 3C). The enrichment pattern differed in soybean, where the clustered genes were enriched for metabolism of nitrogen-containing compounds ($P = 0.037$, hypergeometric test) (Fig. 3C). Clustered genes in sorghum were enriched in terpenoid metabolism ($P = 0.012$, hypergeometric test), whereas phenylpropanoid metabolism was enriched in the

Fig. 3. Clustering and coexpression of specialized metabolic genes.

(A) Distribution of metabolic gene cluster sizes (number of genes per cluster) in four plant genomes. Clusters of size 3 and greater were statistically significant for all four genomes ($P < 2.20 \times 10^{-16}$, χ^2 test) (6). **(B)** Enrichment of functional classes in metabolic gene clusters of *Arabidopsis* and soybean. The x axes represent negative \log_{10} transformations of the P value. **(C)** Enrichment of three main classes of specialized metabolism compounds in clustered versus nonclustered gene sets in *Arabidopsis*, soybean, and sorghum (6). * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$ (hypergeometric test). The y axes represent negative \log_{10} transformations of the P value. **(D)** Coexpression values for clusters containing specialized metabolic gene(s) (red), nonspecialized metabolism-related clusters (blue), known specialized metabolic pathways (black), nonspecialized metabolism-related pathways (gray), randomized clusters (orange), and neighboring genes (green).



nonclustered genes ($P = 0.014$, hypergeometric test) (Fig. 3C). Rice exhibited no enrichment patterns in either gene set. The differing results among the four species suggest that specialized metabolic gene clusters are a product of independent, lineage-specific metabolic evolution, rather than a broad mechanism underlying specialized metabolism evolution.

We used a large-scale microarray data set to test whether the *Arabidopsis* clusters exhibit co-expression (6, 16). Of the 128 clusters with at least three genes, 62 contained at least one specialized metabolism-related gene. The mean co-expression value for the specialized metabolic gene-containing clusters (0.390) was significantly higher than the mean for the remaining clusters (0.289) ($P = 0.020$, Wilcoxon rank sum test) (Fig. 3D). Coexpression of genes within the 62 specialized metabolism-related clusters differed significantly from randomized clusters of genes ($P = 2.71 \times 10^{-7}$, Wilcoxon rank sum test) and neighboring genes in the *Arabidopsis* genome ($P = 3.56 \times 10^{-5}$, Wilcoxon rank sum test) (Fig. 3D) (6). In contrast, coexpression values within the 66 nonspecialized metabolism-related clusters were similar to the examined controls (random clusters, $P = 0.191$; neighboring genes, $P = 0.817$). These results indicate that gene clusters containing specialized metabolic genes are more likely to be coexpressed than their nonspecialized counterparts. Given that genes in the same metabolic pathway

tend to coexpress with each other, newly identified clusters exhibiting high degrees of gene coexpression may represent novel metabolic pathways (fig. S9).

Our findings indicate that the major innovations across plant networks pertain to the emergence of specialized metabolic processes, which, relative to primary metabolic processes, exhibit larger numbers of associated enzymes, increased enzyme proliferation rates, and preferential retention after local gene duplication versus whole-genome duplication. Furthermore, specialized metabolic genes exhibit lineage-specific patterns of genome colocation and specialized metabolism-related gene clusters display heightened levels of gene coexpression in *Arabidopsis*. Collectively, these properties constitute a set of genomic signatures of specialized metabolic genes that may serve as a tool for the accelerated and rational discovery of genes involved in the synthesis of novel specialized compounds (4, 5, 17).

References and Notes

- B. M. Schmidt, D. M. Ribnick, P. E. Lipsky, I. Raskin, *Nat. Chem. Biol.* **3**, 360–366 (2007).
- P. R. Ehrlich, P. H. Raven, *Evolution* **18**, 586–608 (1964).
- M. Wink, *Biochemistry of Plant Secondary Metabolism* (Wiley-Blackwell, Chichester, UK, 2010).
- A. L. Schillmiller, E. Pichersky, R. L. Last, *Curr. Opin. Plant Biol.* **15**, 338–344 (2012).
- J.-K. Weng, R. N. Philippe, J. P. Noel, *Science* **336**, 1667–1670 (2012).
- See supplementary materials on Science Online.

- A.-L. Barabási, Z. N. Oltvai, *Nat. Rev. Genet.* **5**, 101–113 (2004).
- B. C. M. van Wijk, C. J. Stam, A. Daffertshofer, *PLOS ONE* **5**, e13701 (2010).
- B. D. Mishler, *Taxon* **49**, 661–683 (2000).
- J.-K. Weng, J. P. Noel, *Front. Plant Sci.* **4**, 119 (2013).
- D. Nelson, D. Werck-Reichhart, *Plant J.* **66**, 194–211 (2011).
- K. Yonekura-Sakakibara, K. Hanada, *Plant J.* **66**, 182–193 (2011).
- R. C. Moore, M. D. Purugganan, *Curr. Opin. Plant Biol.* **8**, 122–128 (2005).
- D. J. Kliebenstein, *PLOS ONE* **3**, e1838 (2008).
- H. Y. Chu, E. Wegel, A. Osbourn, *Plant J.* **66**, 66–79 (2011).
- T. Obayashi, K. Nishida, K. Kasahara, K. Kinoshita, *Plant Cell Physiol.* **52**, 213–219 (2011).
- K. Yonekura-Sakakibara, K. Saito, *Nat. Prod. Rep.* **26**, 1466–1487 (2009).

Acknowledgments: We thank A. Osbourn, E. Sattely, and K. Dreher for comments. Supported by a Becas Chile-Conicyt postdoctoral fellowship (R.N.-P.) and by NSF grants DBI-0640769 and IOS-1026003. Data are available as supplementary data files S1 to S5. The enzyme annotation program is available for download at <https://dpb.carnegiescience.edu/labs/rhee-lab/software>.

Supplementary Materials

www.sciencemag.org/content/344/6183/510/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S4
Data Files S1 to S5
References (18–51)

11 February 2014; accepted 7 April 2014
10.1126/science.1252076

Deception by Flexible Alarm Mimicry in an African Bird

Tom P. Flower,^{1,2*} Matthew Gribble,² Amanda R. Ridley^{1,3}

Deception is common in nature, but victims of deception discriminate against and ultimately ignore deceptive signals when they are produced too frequently. Flexible variation of signals could allow evasion of such constraints. Fork-tailed drongos (*Dicrurus adsimilis*) use false alarm calls to scare other species away from food that they then steal. We show that drongos mimic the alarms of targeted species. Further, target species reduce their response to false alarm calls when they are repeated. However, the fear response is maintained when the call is varied. Drongos exploit this propensity by changing their alarm-call type when making repeated theft attempts on a particular species. Our results show that drongos can evade the frequency-dependent constraints that typically limit deception payoffs through flexible variation of their alarm calls.

Animal signals tend to be honest (1–3), but deception can persist where it is beneficial for individuals to respond to a signal type on average (4). Such deception stops working when deceptive signals occur too often relative to their honest counterparts (5, 6). This balance

imposes a constraint on payoffs for individuals possessing a single inflexible deceptive signal whose success depends on the abundance of honest models (5). Novel deceptive signals can confer benefits because individuals that produce rare signals are at a selective advantage (7, 8). Nevertheless, individual success still declines as signal frequency increases (7, 8). Some deceivers can tune their signal to better deceive their host (9, 10). However, the benefits available from flexibly alternating between different deceptive signal types remain unknown. We tested whether such flexible variation enables individuals to produce the specific signal most likely to deceive

hosts, and to evade frequency-dependent constraints on deception, by changing their signal when hosts habituate to a previously used signal.

We investigated the deceptive tactics of an African bird, the fork-tailed drongo (*Dicrurus*

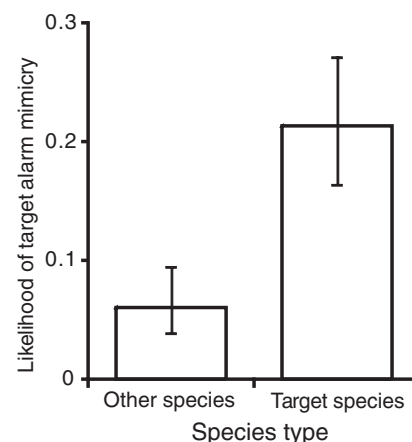


Fig. 1. Mimicry of target species alarm calls in food-theft attempts. Drongos mimicked a target species' alarm calls more often in food-theft attempts on them (target species) than in attempts on other target species (other species) [generalized linear mixed model (GLMM) (binomial), $\chi^2_1 = 2.55$, $P = 0.011$, $n = 147$ food-theft attempts by 10 drongos (table S1)]. Error bars indicate SEM.

¹Department of Science and Technology/National Research Foundation Centre of Excellence at the Percy FitzPatrick Institute, University of Cape Town, Rondebosch 7701, South Africa. ²Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. ³Centre for Evolutionary Biology, School of Animal Biology (M092), University of Western Australia, Crawley 6009, Australia.

*Corresponding author. E-mail: thomas.flower@uct.ac.za