Published in final edited form as:
Science. 2012 August 31; 337(6098): . doi:10.1126/science.1221636.

A gain of function polymorphism controlling complex traits and fitness in nature

Kasavajhala V. S. K. Prasad[1,*], **Bao-Hua Song**[1,*], **Carrie Olson-Manning**[1,*], **Jill T. Anderson**[1], **Cheng-Ruei Lee**[1], **M. Eric Schranz**[1,†], **Aaron J. Windsor**[1,‡], **Maria J. Clauss**[2], **Antonio J. Manzaneda**[1,#], **Ibtehaj Naqvi**[1,§], **Michael Reichelt**[2], **Jonathan Gershenzon**[2], **Sanjeewa G. Rupasinghe**[3,¶], **Mary A. Schuler**[3], and **Thomas Mitchell-Olds**[1,**]

[1]Department of Biology, Duke University, Durham, NC 27708

[2]Max Planck Institute for Chemical Ecology, D-07745 Jena, Germany

[3]Department of Cell and Developmental Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801

Abstract

Identifying the causal genes that control complex trait variation remains challenging, limiting our appreciation of the evolutionary processes that influence polymorphisms in nature. We cloned a QTL that controls plant defensive chemistry, damage by insect herbivores, survival, and reproduction in the natural environments where this polymorphism evolved. These ecological effects are driven by duplications in the *BCMA* loci controlling this QTL and by two selectively favored amino acid changes in the glucosinolate-biosynthetic P450s that they encode. These changes cause a gain of novel enzyme function, modulated by allelic differences in catalytic rate and gene copy number. Ecological interactions in diverse environments likely contribute to the widespread polymorphism of this biochemical function.

Few studies have identified the genes that underlie complex trait variation in nature and the evolutionary processes that influence these polymorphisms. Most work has focused on loss of function mutations that lead to adaptive phenotypes (1), likely because novel gain of function changes occur infrequently and require persistent natural selection to be maintained in populations (2). Nevertheless, new functional mechanisms are crucially important for adaptive evolution (3). To understand the adaptive consequences of complex trait variation we must establish a direct relationship from genetic polymorphisms to phenotypic traits, and investigate their causal effects on fitness in natural environments (1).

**Correspondence to: tmo1@duke.edu.
*These authors contributed equally to the manuscript.
†Present address: Biosystematics Group, Wageningen University, Wageningen, The Netherlands.
‡Present address: Bayer CropScience, N.V., 9052 Gent, Belgium.
#Present address: Departmento de Biologia Animal, Biologia Vegetal y Ecologia, Area de Ecologia, Universidad de Jaen, 23071 Jaen, Spain.
§Present address: Duke University School of Medicine, Durham, NC 27710
¶Present address: Pfizer Inc. MS 8118A-2053, Eastern Point RD, Groton, CT 06340.

Supplementary Materials:
www.sciencemag.org
Materials and Methods
Supporting Online Text
Figs. S1 – S8
Tables S1 – S20
References (22–56)

Glucosinolates (GS) are biologically active secondary compounds (Fig. S1) found in Arabidopsis and its relatives (4) that are important in many aspects of plant defense, influencing oviposition and feeding by insect herbivores (5), defense against microbial pathogens (6), and composition of associated microbial communities (7). Typically, generalist insects are sensitive to glucosinolate-based plant defenses, whereas specialists may be able to cope with these compounds, which may serve as oviposition cues and feeding stimulants (5).

The ecological model plant *Boechera stricta* (Brassicaceae) is a native, short-lived perennial, with a close phylogenetic relationship to Arabidopsis (8), often found in undisturbed habitats where current environments are similar to historical conditions that have existed for ~3,000 years (9). In the field we measured natural selection on herbivore damage using local genotypes. We mapped a QTL in *B. stricta* that contributes to insect resistance and controls allocation to glucosinolates derived from branched chain amino acids or methionine (the Branched Chain Methionine Allocation locus: BCMA,10). Although most Brassicaceae synthesize GS from methionine or tryptophan precursors, among the genera closely related to Arabidopsis, only *Boechera* produces GS from branched chain amino acids (BC-GS), valine or isoleucine (11).

We examined variation in *B. stricta* for three herbivory-related traits (leaf damage by herbivores, total GS, and BC-RATIO, the proportion of GS derived from branched chain amino acids, valine or isoleucine) in nine natural populations in ID and MT, and found significant genetic variation for all traits (Table S1) (12). Levels of herbivore damage (percentage of leaf area removed) and total quantity of foliar glucosinolates showed genetic variation (Fig. 1 and Table S1, $P = 0.0355$ for herbivore damage among families, other $P < 0.0001$) and continuous phenotypic distributions typical of complex traits. In contrast, we found a discontinuous distribution in the proportion of aliphatic glucosinolates derived from branched-chain amino acid precursors versus methionine-derived glucosinolates (Fig. 1), which corresponds to the *BCMA* QTL (13, 14). The parental genotypes examined here have GS phenotypes that are representative of other plants in these populations (14).

We quantified ecological effects of this variation by measuring late-season foliar GS in 1,030 F6 near-isogenic line (NIL) plants in the MT and CO field sites. Segregation of the *BCMA* locus predicted BC-RATIO in both environments ($P < 10^{-199}$, Table S2) and both were significant predictors of insect damage (Table S3, $P < 0.008$). Early-season herbivory showed significant effects of *BCMA* in MT ($P < 10^{-6}$, Table S4) but not in CO. However, the *BCMA* x site interaction for herbivore damage was not significant early in the season ($P = 0.092$, Table S4) or for maximum damage on plants that survived through the summer (Table S3). The quantitative level of leaf damage was a significant predictor of mortality in both MT and CO ($P = 0.0017$; Table S5), with no hint of heterogeneous selection gradients among sites (damage × site interaction, $P = 0.48$). Combining the observed levels of damage and estimates of natural selection (Table S4, Fig. 2), we calculate that the *BCMA-MT* homozygote had 1.3% higher fitness than the *BCMA-CO* genotype (14).

We compared herbivore damage and fecundity on 1,435 Recombinant Inbred Line (RIL) plants, in 2009 in MT and in 2010 in CO. Mean herbivore damage was higher in CO (64.2% ± 5. 8%) than MT (10.3% ± 1.9%). *BCMA* genotype predicted herbivore damage in MT, with the *BCMA-CO* homozygote showing higher damage than the native *BCMA-MT* genotype (14.0% and 9.0% respectively, $P < 0.0001$, Table S6). In contrast, *BCMA* genotypes showed no significant difference in damage levels in CO ($P = 0.63$), perhaps because herbivores in CO are resistant to these compounds, or chemical defenses were overwhelmed by high levels of herbivory. This *BCMA* x site interaction for herbivore damage was significant ($P = 0.019$, Table S6). In addition, herbivore damage was a

significant predictor of fecundity in MT (Fig. 2, P < 0.013, Table S7) but not in CO (P = 0.72), with a significant site x *BCMA* interaction for probability of fruiting (P = 0.05, Table S8). In Montana the local allele enhances the probability of fruiting by 132% relative to the CO allele (14), however, there is no effect of allelic variation in Colorado. Combining observed levels of damage and estimates of natural selection, we calculate that the *BCMA-MT* homozygote had 12% higher fecundity in MT than the CO homozygote (Fig 2, Table S6, 14). Such large fitness differences may explain why many populations are nearly fixed for *BCMA* (Fig. 1B). Overall, the protective effect of *BCMA* against herbivory appears to differ between sites (Table S6), while substantial fitness reduction due to leaf damage is commonly observed across sites and years (Fig. 2).

To control for effects on fitness caused by linked genes, as well as other selective factors that might be correlated with herbivore damage, we planted 1,539 F6 NIL plants where each was assigned to an undamaged control treatment, or to artificial herbivory where we removed ~33% of each leaf (14). On average, a loss of 1% of leaf area caused a 1.2% reduction in survival (Fig. 2), with significantly elevated mortality in the herbivory treatment (P < 0.0001, Table S9).

In *Arabidopsis thaliana*, the *CYP79F* locus encodes the first step of the core Met-glucosinolate pathway (15, 16) and *CYP83A* encodes the second step (17, 18). We sequenced Bacterial Artificial Chromosomes (BACs) from both genotypes that gave rise to the RIL and NIL populations, and identified 9 markers within the 1 cM interval containing the *BCMA* QTL. These included orthologous sequences to *CYP79F* and *CYP83A. CYP83A* is 0.33 cM from the LOD peak for the *BCMA* QTL (Fig. S2; Table S10), while the *CYP79* polymorphism has peak LOD = 365.3, with 10 LOD confidence interval < 0.1 cM wide (Table S10; Fig. S2).

We created transgenic *Arabidopsis* for each locus and allele of the *CYP79* gene family from *Boechera* (14), verifying that *BCMA* belongs to a gene family with three expressed copies (Fig. S3). These enzymes convert amino acids to their corresponding oximes in the first step of glucosinolate biosynthesis (16). *BCMA2* is syntenic with the *CYP79F1* region in *A. thaliana*, but is not linked to the *BCMA* QTL that controls BC-RATIO. *BCMA3* and *BCMA1* are tightly linked at the LOD peak of the *BCMA* QTL, with *BCMA3* present in both parental genotypes, while *BCMA1* is only present in the MT genotype.

We expressed these *BCMA* sequences and controls in 130 independent Arabidopsis transformants to control for position effects and number of insertions. We found significant differences in foliar glucosinolates derived from methionine, valine, or isoleucine in transgenic plants (Figs. 3A & S4; Table S11, $P < 10^{-45}$). *BCMA1-MT* transgenics show increased production of isoleucine-derived glucosinolates ($P < 0.04$), and both *BCMA3* alleles cause increased production of valine-derived glucosinolates ($P = 0.034$ and $P = 0.0001$ for *BCMA-CO* and *BCMA-MT*, respectively) relative to controls. In addition, *BCMA1-MT* and both *BCMA2* alleles cause modest increases in methionine-derived glucosinolates ($P = 0.028$ to $P = 0.002$).

We compared *BCMA3-MT* and *BCMA3-CO* transgenics, and found no significant difference for total concentration of aliphatic glucosinolates (Table S11, $P > 0.05$). However, the *BCMA3-MT* allele produced 3.5-fold higher levels of valine-glucosinolates than the *BCMA3-CO* allele ($P = 0.0002$). The *CO* allele differs from *BCMA3-MT* by an amino acid substitution in the substrate binding region, which may be responsible for this reduced concentration of valine-glucosinolates in transgenic plants. Finally, allele-specific expression to test for cis-regulatory variation found no significant differences in gene expression (14).

Heterologous expression in *E. coli* also indicated that the enzymes encoded by *BCMA1* and *BCMA3*, but not *BCMA2*, have acquired catalytic activity towards branched chain amino acid precursors (the "BC-AA clade," which includes an orthologous sequence from *B. retrofracta*, which also produces BC-GS, 11). Comparing the rate of nonsynonymous versus synonymous substitution with maximum-likelihood in PAML (19) indicated that the branch leading to the BC-AA clade (branch F in Fig. S5) has undergone accelerated biochemical evolution ($P = 0.036$, Table S12). In addition, two amino acid sites (134 and 536) in the BC-AA clade also showed rapid evolution (Table S13).

The BCMA1-MT enzyme has evolved elevated activity towards isoleucine (Fig. 3B, Table S14). For valine, we found significant catalytic activity for BCMA1, BCMA3, and a modest increase for BCMA2. Although transgenic analysis showed that the *BCMA3-MT* allele produced higher levels of valine-glucosinolates than the *BCMA3-CO* allele, heterologous expression did not detect a significant difference in the rate of valine catalysis between these *BCMA3* alleles ($t = 1.09$, df = 6, $P = 0.32$). This may reflect differences in experimental variation between transgenic plants and *in vitro* assays, glucosinolate turnover *in vivo*, or between enzyme function *in vitro* vs. *in vivo*. Finally, we mutated G134 and P536 in BCMA2 (to BCMA1/3 residues L and K, respectively) to assay their effect on catalytic activity on valine and isoleucine. Either mutation, or both together, cause increased activity towards valine (Fig. 3B, Table S14, P = 0.0361 – 0.0004).

The tertiary structures of eukaryotic P450 proteins are highly conserved in spite of substantial divergence in their primary structures (20, 21). We predicted the structure of BCMA2 (Fig. 4) and visualized the locations of variations in the BCMA1 and BCMA3 proteins that might explain their altered catalytic functions. G134L, one of the two residues showing evidence for accelerated molecular evolution, occurs in substrate recognition site SRS1 near the heme (Fig. 4) and is predicted to alter the catalytic site space in the region closest to the heme. The other, P536K, occurs five amino acids upstream from their carboxy-termini and is predicted to alter electrostatic interactions of this flexible tail region. Mapping of the two positions varying between the BCMA3-MT and BCMA3-CO alleles indicates that V148L occurs in a region potentially affecting interactions with electron transfer partners and that M268V occurs in a SRS3 region predicted to affect the volume of the upper catalytic site and/or substrate access (Fig. S6). However, determining the biochemical effects of these changes is beyond the scope of this study.

We have shown here how the *BCMA* QTL affects plant chemistry and insect resistance and thus fitness in a quantitative manner. In *Boechera* the *BCMA2* locus retains ancestral activity and synteny whereas *BCMA1* and *BCMA3* have evolved novel catalytic activity. The resulting polymorphic Met-GS and BC-GS show heterogeneous effects on host plant resistance against diverse enemies across a range of environments. In the Montana population, homozygotes at *BCMA* produce BC-GS and show greater resistance to damage by a diverse community of herbivores (Tables S4, S6). Further evidence that these compounds have environment-dependent consequences comes from transgenic Arabidopsis, where BC-GS cause increased resistance to the pathogen *Erwinia carotovora* (6), and from other herbivores, where BC-GS cause increased susceptibility to *Trichoplusia ni* (10). However, *BCMA* has no effect on insect damage in CO (Tables S4, S6), where other loci control resistance (Table S6). On the basis of this study we conclude that heterogeneous responses to diverse biotic interactions in the context of selection by herbivores likely contributes to the genetic diversity of *BCMA*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank R. Colautti, K. Donohue, M. Feder, T. Pendergast, M. Rausher, C. Rushworth, A. Shumate, M. Wagner, and J. Willis for helpful comments. K. Springer, E. Ballweg, M. Cameron, K. Chu, S. Hurst, V. Cousins, K. Dales, R. Doll, J. Lutkenhaus, M. Mitchell-Olds, S. Mitchell-Olds, E. Raskin, J. Rivera, L. Saucier, M. Wagner, and T. Weiss-Lehman helped in lab and field. N. Wicks and Bitterroot National Forest allowed us to work on their property. We thank the Heald family for support and hospitality. Supported by the National Institutes of Health (award R01-GM086496 to TMO; award R01-GM079530 to MAS) the National Science Foundation (award EF-0723447 to TM-O; dissertation grants 1110445 to CO-M and 1110445 to C-RL), and an NWO-Ecogenomics grant to MES. The data from this paper are in the Supporting Online Material and as Genbank accessions BCMA JQ337904 - BCMA JQ337909.

References and Notes

1. Barrett RD, Hoekstra HE. Molecular spandrels: tests of adaptation at the genetic level. Nat Rev Genet. 2011; 12:767. [PubMed: 22005986]
2. Rogers RL, Hartl DL. Chimeric genes as a source of rapid evolution in *Drosophila melanogaster*. Mol Biol Evol. 2012; 29:517. [PubMed: 21771717]
3. Futuyma DJ, Agrawal AA. Macroevolution and the biological diversity of plants and herbivores. Proc Natl Acad Sci USA. 2009; 106:18054. [PubMed: 19815508]
4. Halkier BA, Gershenzon J. Biology and biochemistry of glucosinolates. Ann Rev Plant Biol. 2006; 57:303. [PubMed: 16669764]
5. Hopkins RJ, van Dam NM, van Loon JJA. Role of glucosinolates in insect-plant relationships and multitrophic interactions. Ann Rev Entomol. 2009; 54:57. [PubMed: 18811249]
6. Brader G, Mikkelsen MD, Halkier BA, Palva ET. Altering glucosinolate profiles modulates disease resistance in plants. Plant J. 2006; 46:758. [PubMed: 16709192]
7. Bressan M, et al. Exogenous glucosinolate produced by *Arabidopsis thaliana* has an impact on microbes in the rhizosphere and plant roots. ISME Journal. 2009; 3:1243. [PubMed: 19554039]
8. Rushworth C, Song B, Lee CR, Mitchell-Olds T. *Boechera*, a model system for ecological genomics. Molec Ecol. 2011; 20:4843. [PubMed: 22059452]
9. Brunelle A, Whitlock C, Bartlein P, Kipfmueller K. Holocene fire and vegetation along environmental gradients in the Northern Rocky Mountains. Quat Sci Rev. 2005; 24:2281.
10. Schranz ME, Manzaneda AJ, Windsor AJ, Clauss MJ, Mitchell-Olds T. Ecological genomics of *Boechera stricta*: identification of a QTL controlling the allocation of methionine- vs branched-chain amino acid-derived glucosinolates and levels of insect herbivory. Heredity. 2009; 102:465. [PubMed: 19240753]
11. Windsor AJ, et al. Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (Brassicaceae). Phytochemistry. 2005; 66:1321. [PubMed: 15913672]
12. Materials and methods are available as supplementary material in Science Online.
13. Schranz ME, Windsor AJ, Song B-h, Lawton-Rauh A, Mitchell-Olds T. Comparative genetic mapping in *Boechera stricta*, a close relative of *Arabidopsis*. Plant Physiol. 2007; 144:286. [PubMed: 17369426]
14. See Supporting Online Text in Science Online.
15. Reintanz B, et al. Bus, a bushy arabidopsis CYP79F1 knockout mutant with abolished synthesis of short-chain aliphatic glucosinolates. Plant Cell. 2001; 13:351. [PubMed: 11226190]
16. Chen SX, et al. CYP79F1 and CYP79F2 have distinct functions in the biosynthesis of aliphatic glucosinolates in *Arabidopsis*. Plant J. 2003; 33:923. [PubMed: 12609033]
17. Bak S, Tax FE, Feldmann KA, Galbraith DW, Feyereisen R. CYP83B1, a cytochrome P450 at the metabolic branch paint in auxin and indole glucosinolate biosynthesis in Arabidopsis. Plant Cell. 2001; 13:101. [PubMed: 11158532]
18. Naur P, et al. CYP83A1 and CYP83B1, two nonredundant cytochrome P450 enzymes metabolizing oximes in the biosynthesis of glucosinolates in *Arabidopsis*. Plant Physiol. 2003; 133:63. [PubMed: 12970475]

19. Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007; 24:1586. [PubMed: 17483113]

20. Poulos, TL.; Johnson, EF. Cytochrome P450: Structure, Mechanism, and Biochemistry. Ortiz de Montellano, PR., editor. Kluwer Academic / Plenum Publishers; New York: 2005.

21. Poulos, TL.; Meharenna, YT. The Ubiquitous Roles of Cytochrome P450 Proteins. Sigel, A.; Sigel, H.; Sigel, RKO., editors. John Wiley & Sons, Ltd; Chichester, UK: 2007. p. 57-96.

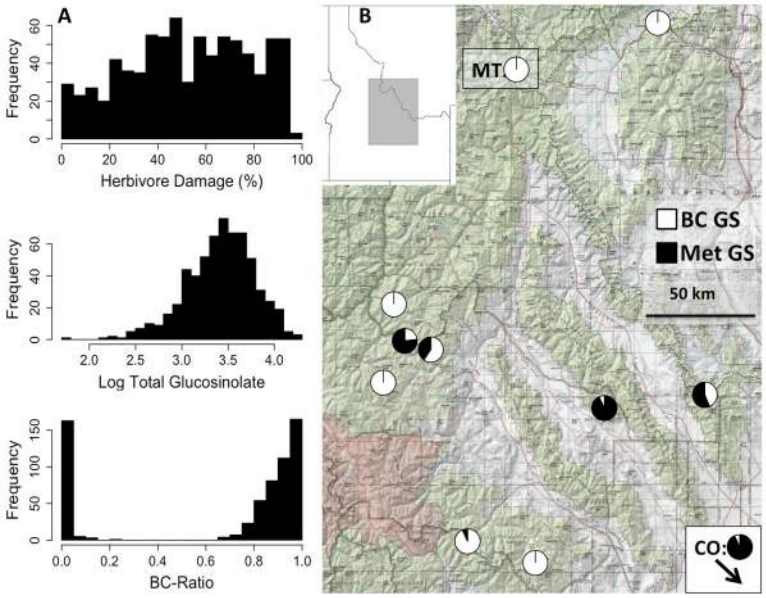

Fig. 1.

(A) Histograms show percent leaf area removed by the generalist herbivore *Trichoplusia ni*, total quantity of glucosinolates, and proportion of aliphatic glucosinolates from branched chain amino acid precursors. Greenhouse-grown plants descended from nine *B. stricta* populations. (B) Map showing the proportion of genotypes in each population that produce predominantly branched-chain derived glucosinolates (white) or methionine-derived glucosinolates (black). Parental populations of the crossing experiment are boxed.

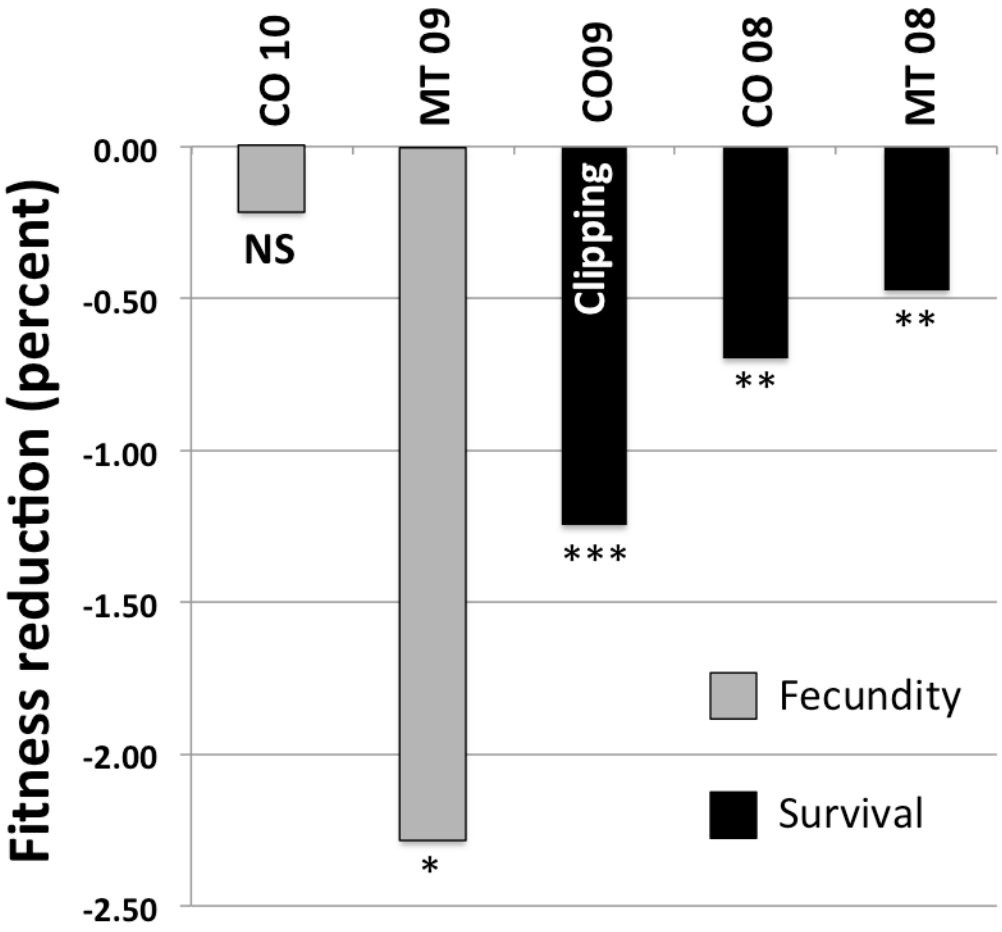

Fig. 2.

Fitness reductions under field conditions associated with 1% loss of leaf area by herbivory. Bars indicate reduction in components of fitness due to fecundity (gray) and survival (black) in 2008, 2009, and 2010 in Colorado and Montana. NIL plants in the clipping experiment were randomly assigned to artificial herbivory or control treatments. * = P < 0.05, ** = P < 0.01, *** = P < 0.001, NS = not significant.

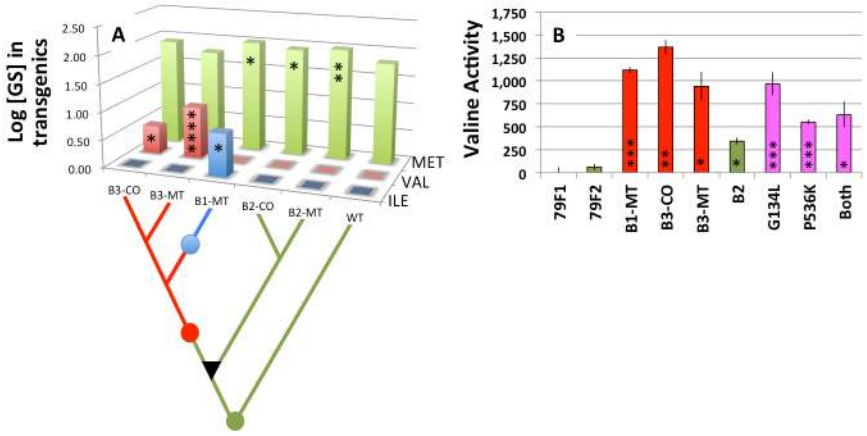

Fig. 3.

(A) Glucosinolate production in transgenic Arabidopsis expressing *B. stricta BCMA* genes, encoding CYP79F enzymes which catalyze amino acids in the first step of the glucosinolate pathway. Bars show amounts of aliphatic glucosinolates from MET, VAL, and ILE precursors catalyzed by each of the *BCMA* gene products. Gene phylogeny includes wildtype (WT) Arabidopsis with empty vector controls, the black triangle identifies the gene duplication in *Boechera*, red circle shows origin of branched chain amino acid catalysis, and blue circle indicates elevated ILE activity. Abbreviations: B1 = *BCMA1*, B2 = *BCMA2*, B3 = *BCMA3*, with alleles from CO or MT. N = 130 independent transgenic lines. (B) *In vitro* enzyme activity levels relative to controls, in nmol product/nmol enzyme/min, with standard error. Labels indicate CYP79F enzymes from Arabidopsis, and BCMA1, BCMA2, and BCMA3 from *Boechera*, with alleles from CO or MT. BCMA2 alleles encode identical proteins, so one allele was assayed. BCMA2 (green) retains the ancestral MET activity, and was engineered to change G134L, P536K, or both (pink). * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$, **** = $P < 0.0001$.

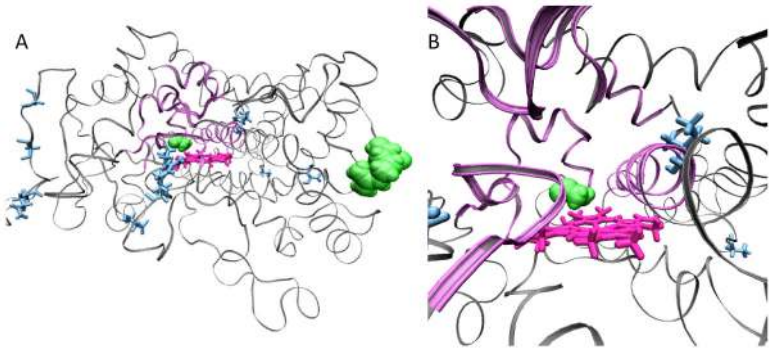

Fig. 4.

(A) Homology model of BCMA2 with the substrate binding cleft above the heme group (magenta) with putative substrate recognition regions in purple. Amino acid changes G134L and P536K (green) show statistical evidence for accelerated protein evolution, and alter catalytic function when changed by site-directed mutagenesis. Other mutations with statistical evidence of accelerated evolution (in blue) are not addressed in this study. The location of amino acid 529, which aligns with the last resolved residue in the CYP1A2 crystal structure, is colored since subsequent amino acids 530–540 cannot be accurately modeled. (B) Close-up view of substrate binding cleft with mutation G134L residing just above the heme.