WEB346

Benjamin Weigel

July 27th, 2015

Question

Do Mg^{2+}, pH, chemical motif (e.g. catecholic, phenolic, 3'-OMe, 4'-OMe) and the choice of enzyme (WT or 4'-variant) influence the observed conversion of flavonoids and phenylpropanoids by the O-methyltransferase PFOMT?

Introduction

17 different flavonoid and phenyl propanoid substrates were tested for methylation. These substrates can loosely be categorized into four groups by the chemical motif that is to be methylated (e.g. catecholic, phenolic, 3'-OMe, 4'-OMe). Three other factors are studied that might also influence the conversion. The addition of Mg^{2+}, high (8.6) or low (7.5) pH and the enzyme variant (WT or 3'-variant) that is used are also varied.

A TOTAL OF 96 EXPERIMENTS were conducted to cover each group at least with three independent observations.

$$\text{Mg} \times \text{pH} \times \text{variant} \times \text{motif}$$
$$2 \times 2 \times 2 \times 4 = 32$$

Figure 1: number of groups studied

Results

Data

	substrate	value	pH	Mg	variant	motif
1	1	0.00	low	FALSE	WT	A
2	1	0.00	high	FALSE	Y51RN202W	A
3	1	0.07	high	TRUE	WT	A
4	1	0.00	low	TRUE	Y51RN202W	A
5	1	0.04	high	FALSE	WT	A
6	1	0.00	low	FALSE	Y51RN202W	A

Table 1: First rows of dataframe

	labels	motif
1	A	phenolic
2	B	catecholic
3	C	4-O-Me
4	D	3-O-Me

Table 2: Labels in the data.frame and their corresponding motif.

Significance no enzyme vs. enzyme

Do the amounts of produced product vary significantly between treatments were no enzyme was added and trreatm,ents with enzyme? Only ran blanks at high pH with no magnesium added. Subset data to only include high pH, no Mg, no catechols. The catechols

were removed to avoid the bias associated with an autocorrelation between product and catecholic moiety.

From the ANOVA table it is clear, that conversion is not due to chance. The p-value for the variant factor is almost 0. This means, that we should be able to use the data for analyis.

```
lm(value ~ variant * motif, data = data.blank %>%
    filter(pH == "high", Mg == F, motif != "B")) %>%
    aov() %>% summary() %>% xtable(., caption = "ANOVA table for comparison of blanks and samples. There i
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
variant	2	0.01	0.01	66.03	0.0000
motif	2	0.00	0.00	7.72	0.0038
variant:motif	4	0.00	0.00	7.72	0.0008
Residuals	18	0.00	0.00		

Table 3: ANOVA table for comparison of blanks and samples. There is a stitistical significance between treatments, where 0 enzyme was added and treatments that contained enzymne (the wt).

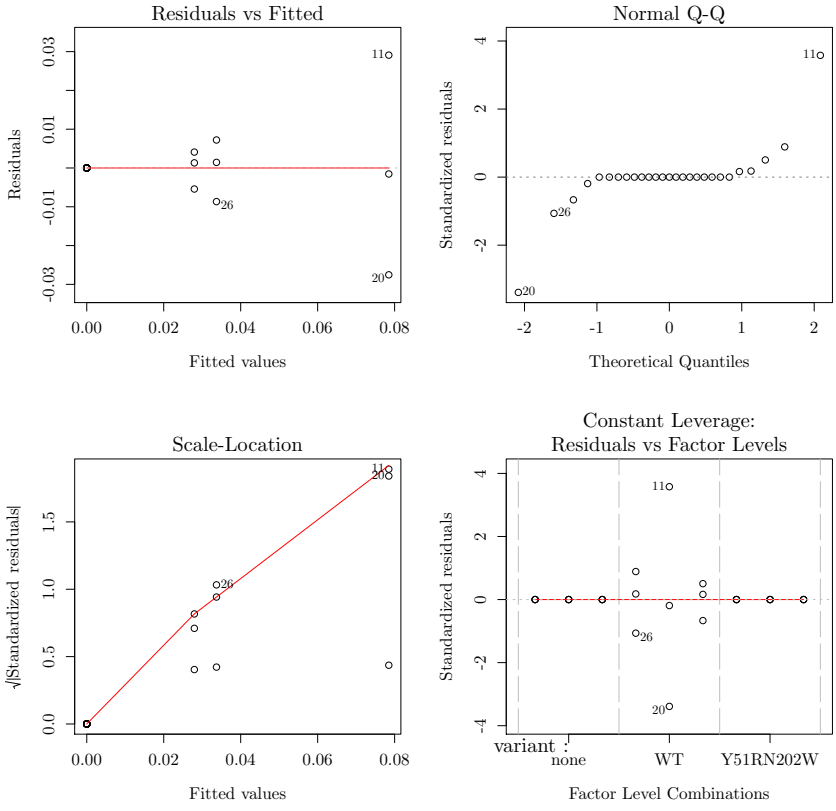


Figure 2: Diagnostic plots of the ANOVA

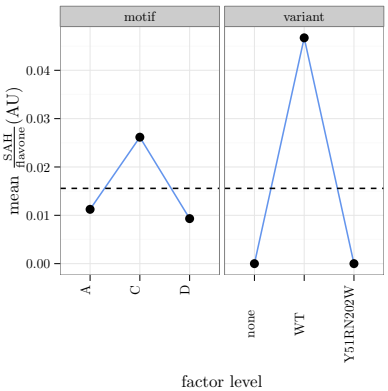


Figure 3: Main effects plot for factors that influence product amount.

Main effects plots

The main effects plots (**??**) give an overview of what happens. The motif clearly has the biggest influence on conversion. This makes sense, since PFOMT acts on catecholic moities.

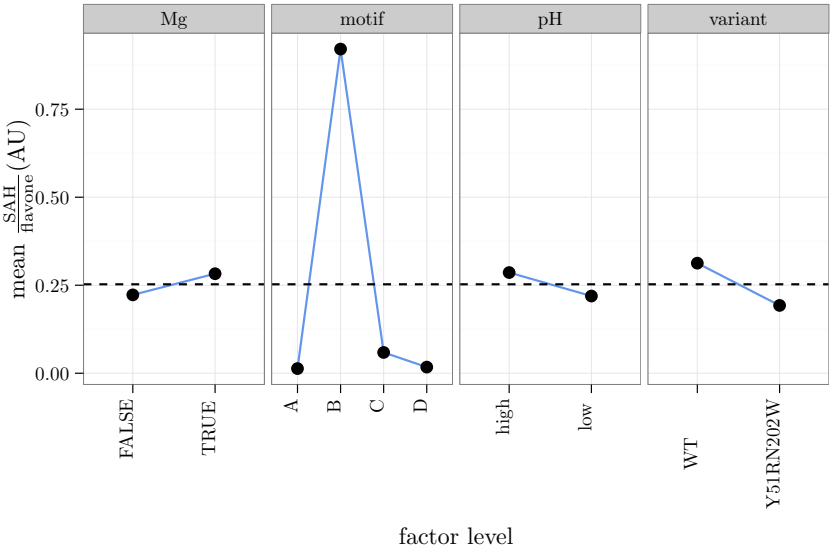


Figure 4: Main effetcs plots for the factor variables. Clearly the motif seems to have the biggest impact. Catecholic moieties are converted most effectively.

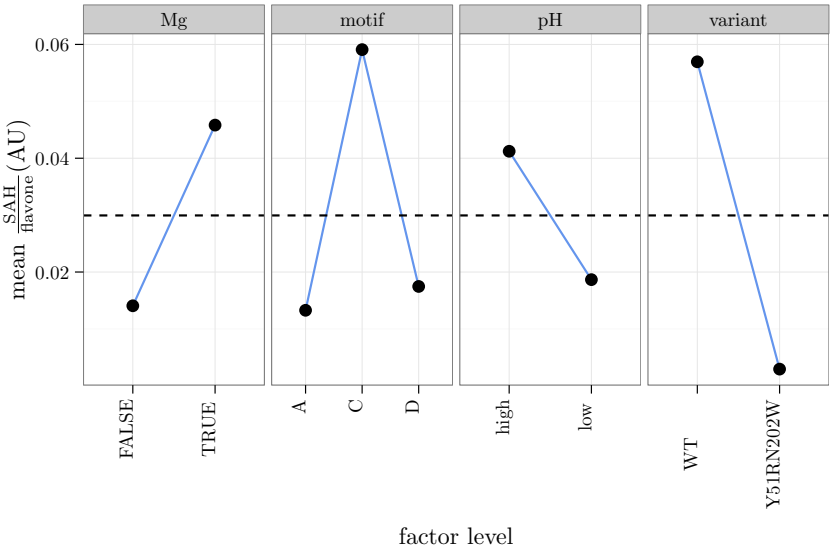


Figure 5: Main effects plot, when catecholic moieties factor is omitted. Phenolics are converted worst. The wildtype has the best activity.

The interaction plot(**??**) of Mg and pH displays an interaction. When magnesium is added the activity tends to be higher. This effect is more pronounced at low pH values. It is not possible to say wether

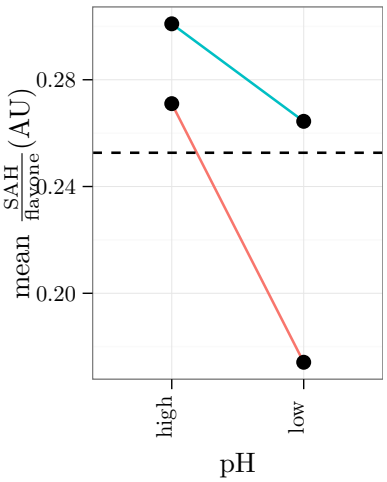


Figure 6: Interaction plot for Mg and pH. The lines suggest an interaction between pH and Mg, but this is not enough evidence to say wether that interaction is significant. red – no magnesium, blue – magnesium added.

this effect is significant without the according satistical test. Indeed the ANOVA results suggest, that there is no significance. In fact there doesn't even seem to be a statistical significance from Mg addition or pH alone.

Anova

Two-way ANOVA test for simple models and more complex models were prepared. The p-values for the complex model are very low all over. However it is very likely that this is an overinterpretation of the data. The data/experiment might be too complex to derive anything of value from the information. The use of more complex modelling techniques mnight shed some light. However, when the substrates with catecholic moieties are excluded from the results, there is statistical significance at leas for Mg and pH.

```
data.lm <- lm(value ~ pH * Mg, data = WEB346.results)
summary(aov(data.lm)) %>% xtable(., caption = "ANOVA table.")
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pH	1	0.11	0.11	0.61	0.4370
Mg	1	0.09	0.09	0.49	0.4837
pH:Mg	1	0.02	0.02	0.12	0.7247
Residuals	92	16.13	0.18		

Table 4: ANOVA table.

```
data.lm <- lm(value ~ pH * Mg * motif * variant,
    data = WEB346.results)
summary(aov(data.lm)) %>% xtable(., caption = "ANOVA table of the most complex model. Significance is ever
```

```
data.lm <- lm(value ~ pH * Mg + variant, data = WEB346.results %>%
    filter(motif != "B"))
summary(aov(data.lm)) %>% xtable(., caption = "ANOVA table when catecholics are excluded. At least the dat
```

Regression Tree

A regression tree can be built from the data. At first glance it shows, that the motif is especially important for conversion. This is trivial, since PFOMT is a 4'-OMT that acts on catecholic motifs. The substrates with catecholic motifs are thus converted more efficiently.

It seems as if the variant is influenced by pH and metal addition more clearly thatn the wt-enzyme, since the tree splits up more.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pH	1	0.11	0.11	24.75	0.0000
Mg	1	0.09	0.09	20.08	0.0000
motif	3	14.31	4.77	1105.42	0.0000
variant	1	0.35	0.35	80.19	0.0000
pH:Mg	1	0.02	0.02	5.07	0.0278
pH:motif	3	0.14	0.05	10.88	0.0000
Mg:motif	3	0.07	0.02	5.24	0.0027
pH:variant	1	0.03	0.03	7.14	0.0096
Mg:variant	1	0.02	0.02	3.52	0.0652
motif:variant	3	0.35	0.12	26.74	0.0000
pH:Mg:motif	3	0.08	0.03	6.49	0.0007
pH:Mg:variant	1	0.02	0.02	4.85	0.0313
pH:motif:variant	3	0.23	0.08	17.67	0.0000
Mg:motif:variant	3	0.20	0.07	15.29	0.0000
pH:Mg:motif:variant	3	0.06	0.02	4.25	0.0084
Residuals	64	0.28	0.00		

Table 5: ANOVA table of the most complex model. Significance is everywhere ... :O

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pH	1	0.01	0.01	4.55	0.0366
Mg	1	0.02	0.02	9.01	0.0038
variant	1	0.05	0.05	26.07	0.0000
pH:Mg	1	0.00	0.00	0.14	0.7082
Residuals	67	0.13	0.00		

Table 6: ANOVA table when catecholics are excluded. At least the data suggest significance for pH and Mg.

Linear Models

At first it was checked, if pH and Mg have an influence on the conversion. It does not seem that Mg or pH have any influence on the conversion, as the p-values are much too high. This could be the result of the fact that the conversion reactions were incubated for 16 hours. Possible intricacies in the conversion (due to different reaction velocities) can not be distinguished from one another if the reaction time is so long that all substrate is used up.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.2711	0.0855	3.17	0.0021
pHlow	-0.0969	0.1209	-0.80	0.4247
MgTRUE	0.0299	0.1209	0.25	0.8051
pHlow:MgTRUE	0.0604	0.1709	0.35	0.7247

Table 7: A linear model with pH and Mg as factors is not a lot better than random guesses. p-values are very high. Thus this alone does not explain the variance.

However, when the catecholics are omitted as a factor level

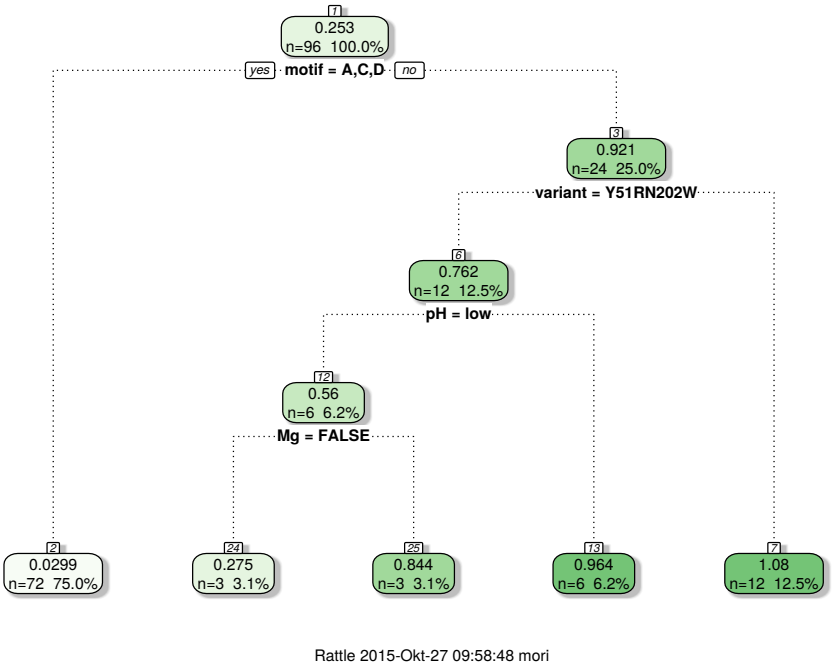


Figure 7: RegressionTree of the data built with the 'rpart'-package. The motif effects the conversion most. It seems that only the variant is influenced by pH and Mg.

it becomes clear that Mg and pH DO in fact influence the conversion. After backwards selection only the factors pH, Mg, variant and te interactions between variant and Mg and pH are retained.

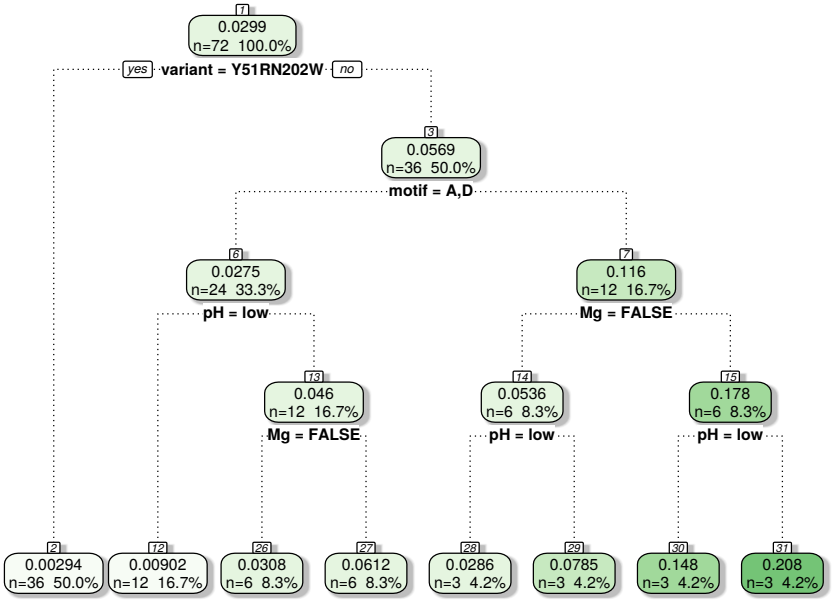
```
data.lm <- lm(value ~ pH * Mg * variant, data = WEB346.results %>%
    filter(motif != "B"))
data.lm <- MASS::stepAIC(data.lm, direction = "backward")
```

```
## Start:  AIC=-447.48
## value ~ pH * Mg * variant
##
##                 Df  Sum of Sq     RSS
## - pH:Mg:variant  1 6.5387e-05 0.11532
## <none>                        0.11526
##                     AIC
## - pH:Mg:variant -449.44
## <none>          -447.48
##
## Step:  AIC=-449.44
## value ~ pH + Mg + variant + pH:Mg + pH:variant + Mg:variant
##
##              Df Sum of Sq     RSS     AIC
## - pH:Mg       1 0.0002845 0.11561 -451.26
## <none>                    0.11532 -449.44
## - pH:variant  1 0.0075587 0.12288 -446.87
```

```
## - Mg:variant  1 0.0120562 0.12738 -444.28
##
## Step:  AIC=-451.26
## value ~ pH + Mg + variant + pH:variant + Mg:variant
##
##              Df Sum of Sq     RSS     AIC
## <none>                    0.11561 -451.26
## - pH:variant  1 0.0075587 0.12316 -448.70
## - Mg:variant  1 0.0120562 0.12766 -446.12
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
pH	1	0.01	0.01	5.23	0.0254
Mg	1	0.02	0.02	10.36	0.0020
variant	1	0.05	0.05	29.98	0.0000
pH:variant	1	0.01	0.01	4.32	0.0417
Mg:variant	1	0.01	0.01	6.88	0.0108
Residuals	66	0.12	0.00		

Table 8: When catecholics are omitted there is significance in the following terms: pH, Mg, variant, pH:variant, Mg:variant



This is a margin note. Notice that there isn't a number preceding the note.

Figure 8: RegressionTree of the data of only substrates with phenolic, 3'-OMe and 4'-OMe moieties. Catecholic substrates are omitted.

Model selection using lasso regression

Lasso regression can shrink model coefficients to zero. This helps in variable selection. Variables that were shrunken to 0 tend to have little impact on the prediction capabilities of the model.

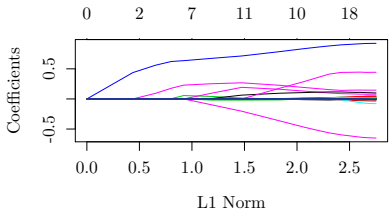


Figure 9: Lasso regression on the model.

Cross-validation to select best model

After cross-validation only 11 variuables have non-zero coefficients and are thus used during prediction. All other variabkes were shrunken to zero.

```
cv.lasso <- cv.glmnet(x, y, alpha = 1, nfolds = 5)
```

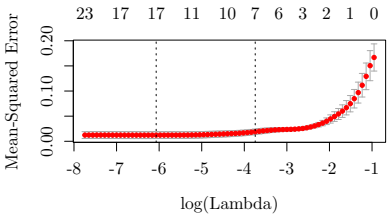


Figure 10: Cross validation results for lasso regression. The best model only needs around 10 variables to describe the data with low error.

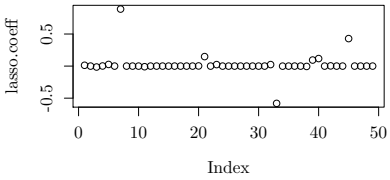


Figure 11: The variables that have non-zero coefficients.

Answer

- the conversion is not by chance (significance between blank and enzyme treated groups)
- The biggest influence on conversion has the presence of a catechgolic moiety (surprise, surprise !)
- There is an effect of Mg and pH (main and interaction effect). However the statistical test don't support this notion. → incubation time too large, this makes results badly interpretable !!! → when catecholic moieties are excluded there is significance for Mg addition and possibly for pH; however no interaction
- data is also very complex with possibly up to four factor interactions
- regression tree used for interpretation → pH and Mg can explain some of the variance

	variable	coefficient
1	(Intercept)	0.0128
2	pHlow	-0.0144
3	MgTRUE	0.0017
4	variantWT	0.0258
5	motifB	0.8890
6	pHlow:variantWT	-0.0096
7	MgTRUE:variantWT	0.0003
8	MgTRUE:motifD	0.0023
9	variantWT:motifB	0.1481
10	variantWT:motifC	0.0234
11	pHlow:variantWT:motifB	0.0237
12	pHlow:variantY51RN202W:motifB	-0.5812
13	pHlow:variantWT:motifC	-0.0005
14	MgTRUE:variantWT:motifB	-0.0052
15	MgTRUE:variantY51RN202W:motifB	0.0944
16	MgTRUE:variantWT:motifC	0.1173
17	MgTRUE:variantWT:motifD	0.0031
18	pHlow:MgTRUE:variantY51RN202W:motifB	0.4286

Table 9: Variables and coefficients that were retained. Non-zero coefficients not shown.

Conversion of substrates

The conversion of the substrate in % is of interest.

Calculation of SAH and SAM concentration

The SAH and SAM concentration were estimated from the area-under-curve (AUC) of the SAM and SAH peaks. The displayed formula also already include the conversion. x_{SAH} is a direct measure for the conversion.

0.4 mM substrate and 0.5 mM biologically active SAM were added. That means the substrate conversion can be estimated by $conversion = \frac{0.4\text{mM}}{0.5\text{mM} \times x_{\text{SAH}}}$.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.9123E-01	2.8871E-01	2.0479E+00	4.2304E-02
area	1.4674E-04	2.6532E-07	5.5305E+02	1.4152E-251

$$A_{\text{SAM}} + A_{\text{SAM}} = 1 \approx 500\text{uM}$$
$$x_{\text{SAH}} = \frac{A_{\text{SAH}}}{A_{\text{SAM}} + A_{\text{SAM}}}$$
$$c_{\text{SAH}} = x_{\text{SAH}} \times 500\text{uM}$$

Figure 12: Calculation of specific activity.

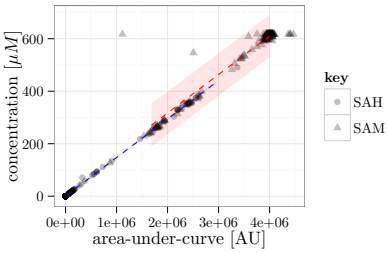


Figure 13: Estimated SAH and SAM concentration plotted against the AUC. Linear best-fit models with 95% prediction intervals are included.

Calculation of the coefficients for concentration~area bay bootstrapping

From the histogram of the coefficients obtained by the bootstrap it is clear that there is a non-gaussian distribution for the coefficients for SAM. This could be due to the fact that there is a huge amount of samples with high SAM concentrations, which produces bias.

```
boot.fn <- function(data, index) {
    data <- data[index, ]
    data %<>% mutate(x = (area/flavon)/((SAH/flavon) +
        (SAM/flavon)), concentration = (x * 500/0.81))  # the total concentration of SAm is needed (R+L) -
    return(coef(lm(concentration ~ area, data = data)))
}


lm.boot.sah <- boot(WEB346.CV %>% select(SAH,
    SAM, flavon) %>% mutate(area = SAH), boot.fn,
    1000)
lm.boot.sam <- boot(WEB346.CV %>% select(SAH,
    SAM, flavon) %>% mutate(area = SAM), boot.fn,
    1000)

pred.SAH <- data.frame(area = WEB346.CV$SAH, concentration = WEB346.CV$SAH *
    lm.boot.sah$t0[2] + lm.boot.sah$t0[1])
pred.SAM <- data.frame(area = WEB346.CV$SAM, concentration = WEB346.CV$SAM *
    lm.boot.sam$t0[2] + lm.boot.sam$t0[1])
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = WEB346.CV %>% select(SAH, SAM, flavon) %>% mutate(area = SAH),
##     statistic = boot.fn, R = 1000)
##
##
## Bootstrap Statistics :
##         original       bias     std. error
## t1* 0.5912346000 4.456157e-03 1.974150e-01
## t2* 0.0001467374 5.576398e-09 3.876305e-07
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
```

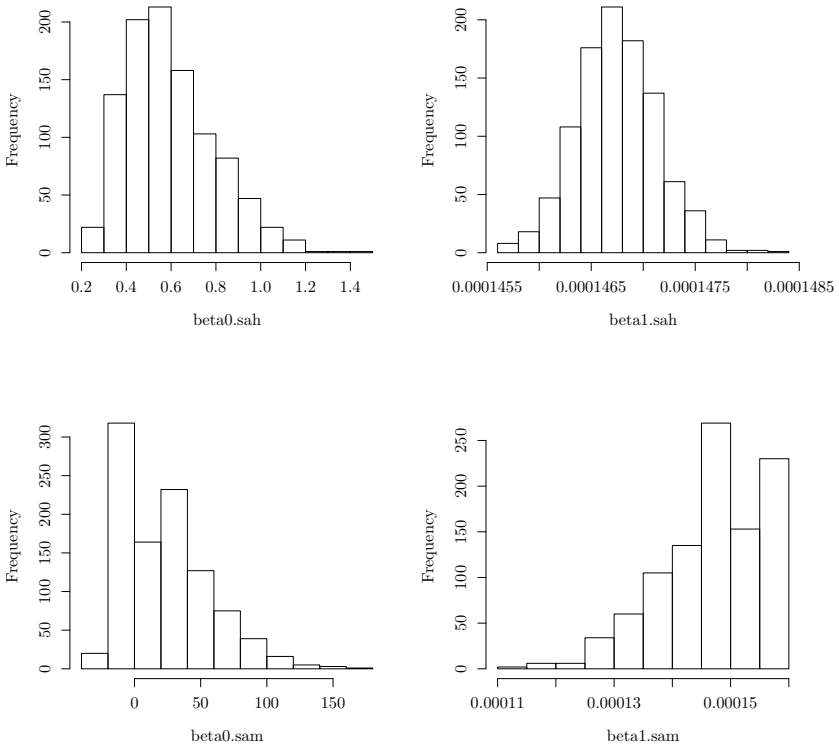


Figure 14: Histogram of coefficient values obtained by bootstrap with n=1000

```
## Call:
## boot(data = WEB346.CV %>% select(SAH, SAM, flavon) %>% mutate(area = SAM),
##     statistic = boot.fn, R = 1000)
##
##
## Bootstrap Statistics :
##         original        bias     std. error
## t1* 2.405931e+01 -1.042784e+00 3.324905e+01
## t2* 1.465407e-04  2.895498e-07 8.687305e-06
```

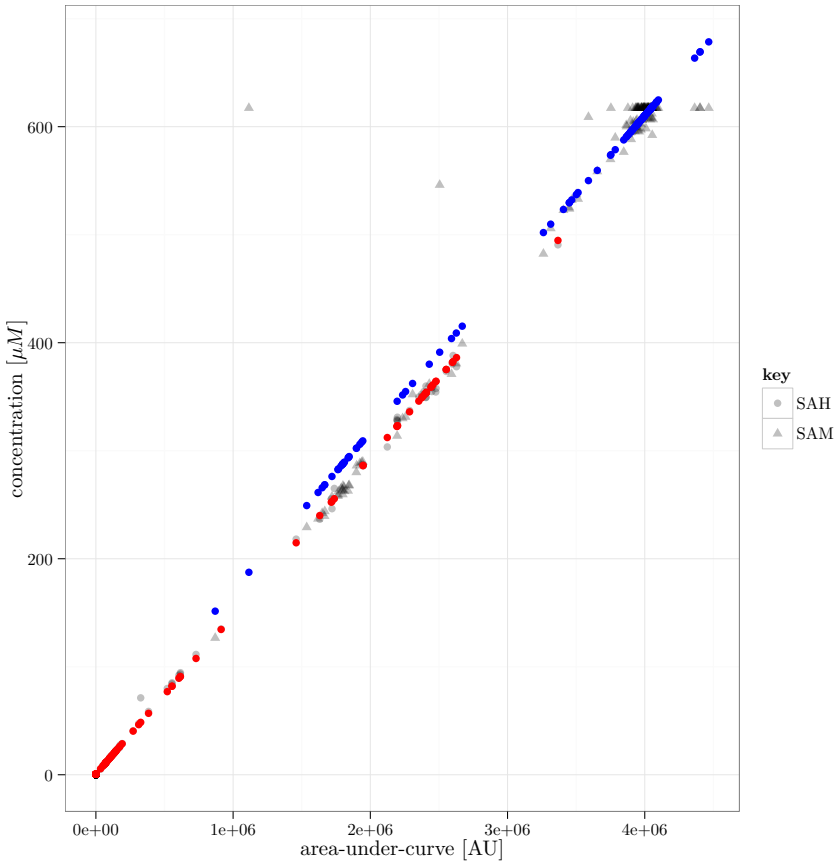


Figure 15: Calculated data from the bootstrap.

Phenolic substrates

3′-O-methyl substrates

4′-O-methyl substrates

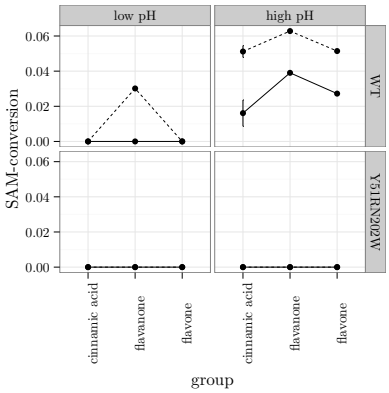


Figure 16: Comparison of conversion of phenolic substrates. dashed line – 10 mM Mg, solid line – no Mg

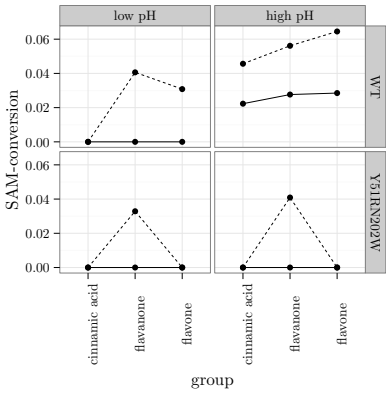


Figure 17: Comparison of conversion of 3′-O-methyl substrates. dashed line – 10 mM Mg, solid line – no Mg

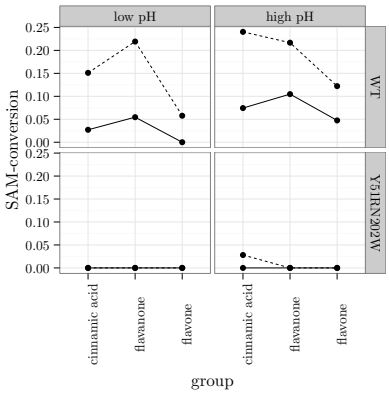


Figure 18: Comparison of conversion of 4′-O-methyl substrates. dashed line – 10 mM Mg, solid line – no Mg

catecholic substrates

Calculation of all conversions

	substrate	variant	max	at.pH	at.Mg
1	1	WT	0.06	high	TRUE
2	2	WT	0.94	high	TRUE
3	3	WT	0.22	low	TRUE
4	4	WT	0.06	high	TRUE
5	5	WT	0.05	high	TRUE
6	6	WT	0.95	low	TRUE
7	7	WT	0.12	high	TRUE
8	8	WT	0.07	high	TRUE
9	9	WT	0.06	high	TRUE
10	10	WT	0.06	high	TRUE
11	11	WT	0.04	high	TRUE
12	12	WT	1.00	low	FALSE
13	14	WT	0.25	high	TRUE
14	13	WT	0.05	high	TRUE
15	15	WT	0.06	high	TRUE
16	16	WT	0.93	high	FALSE
17	17	WT	1.29	low	FALSE

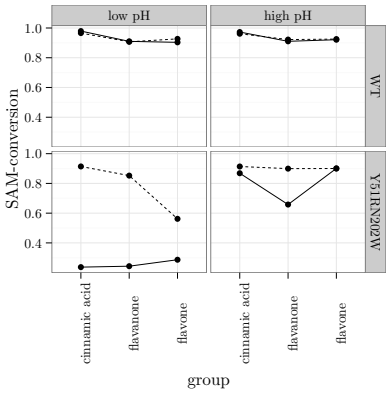


Figure 19: Comparison of conversion of catecholic substrates. dashed line – 10 mM Mg, solid line – no Mg

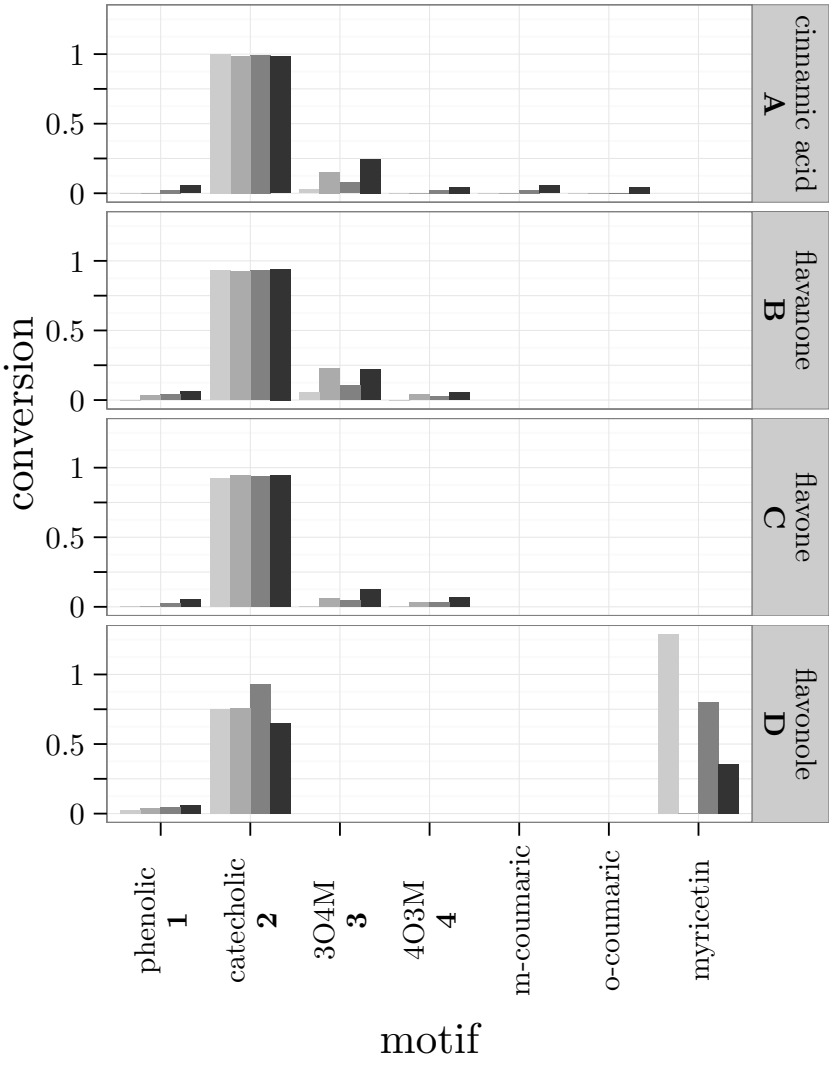


Figure 20: Wild-type conversions. colors from light to dark: low/no, low/yes, high/no, high/yes

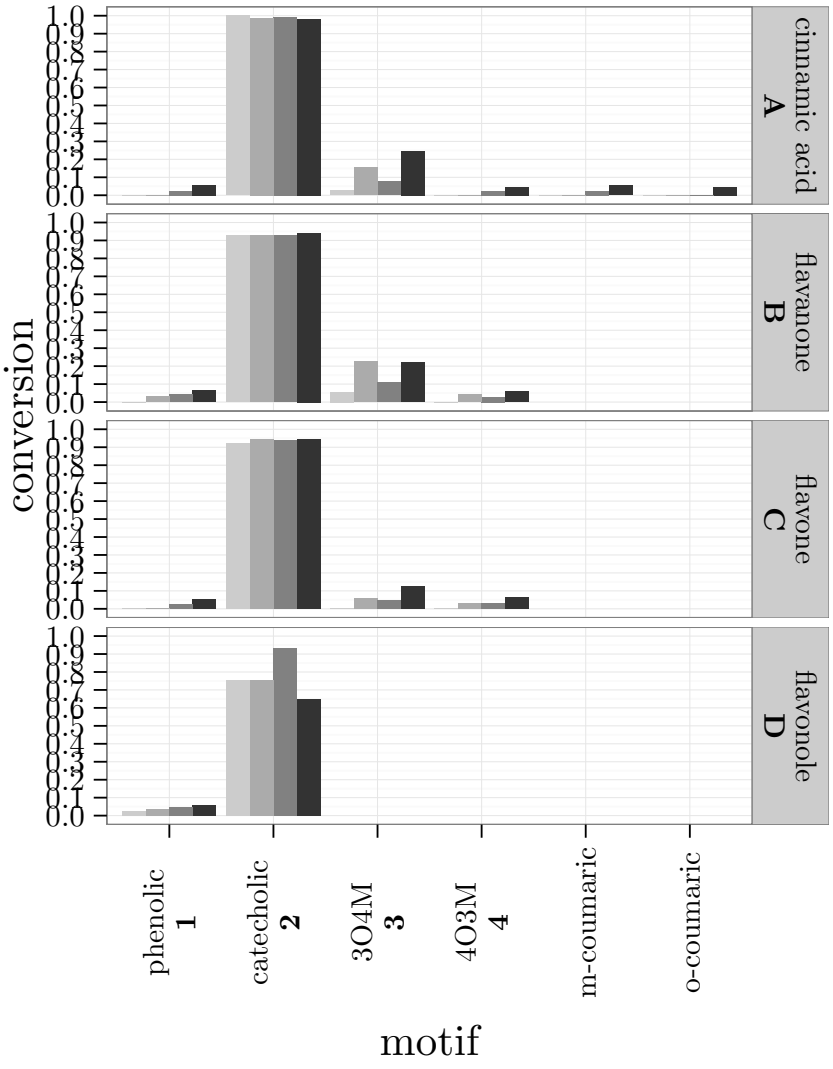


Figure 21: Same as above, but some data is omitted.

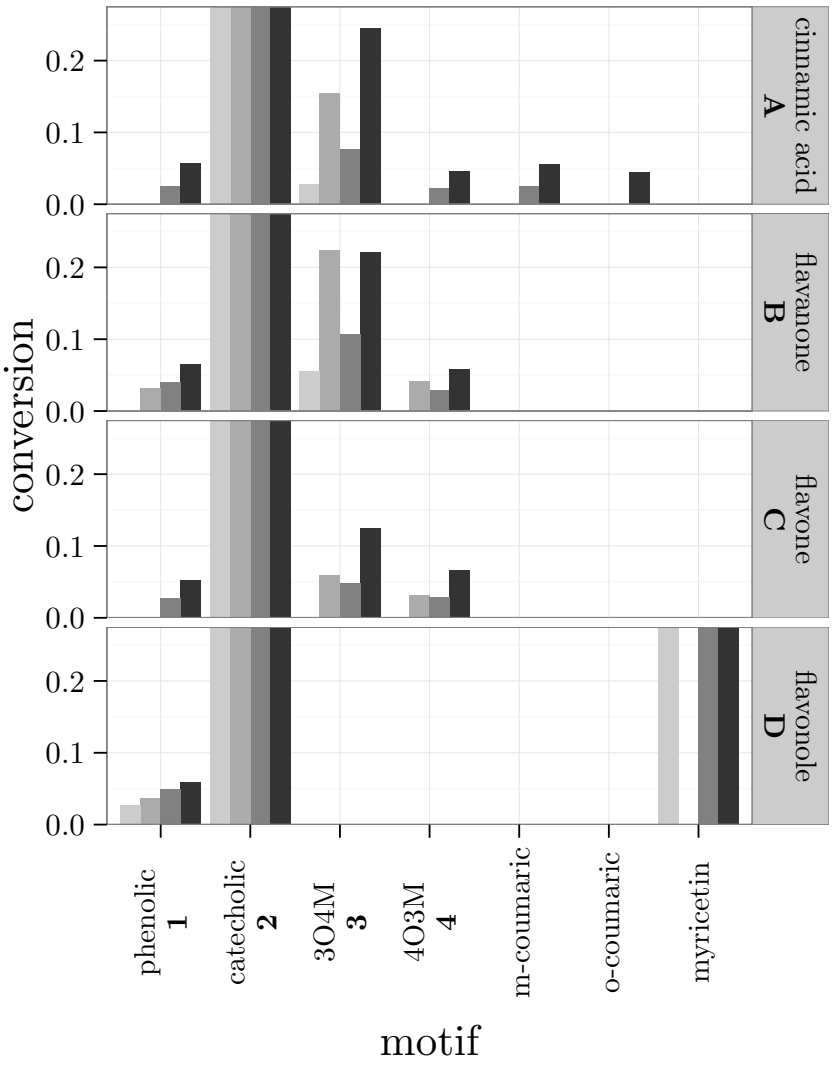


Figure 22: 4′-selective variant conversions. colors from light to dark: low/no, low/yes, high/no, high/yes

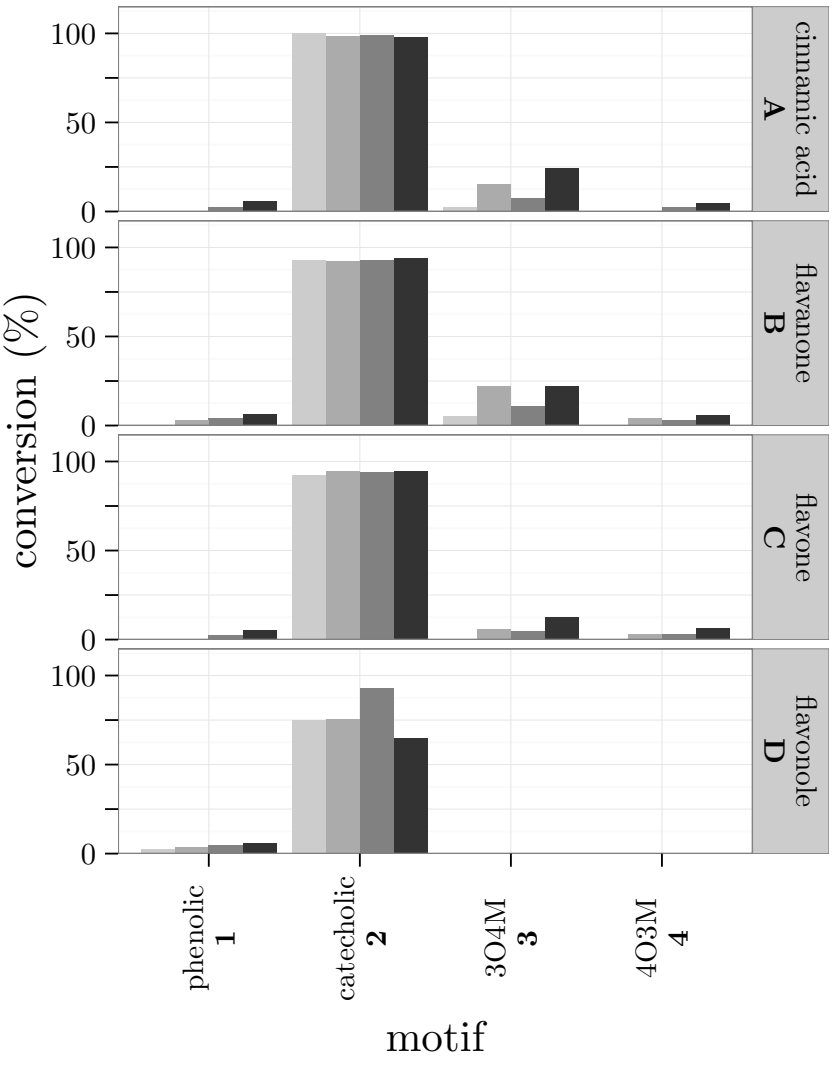


Figure 23: Wild-type conversions. colors from light to dark: low/no, low/yes, high/no, high/yes

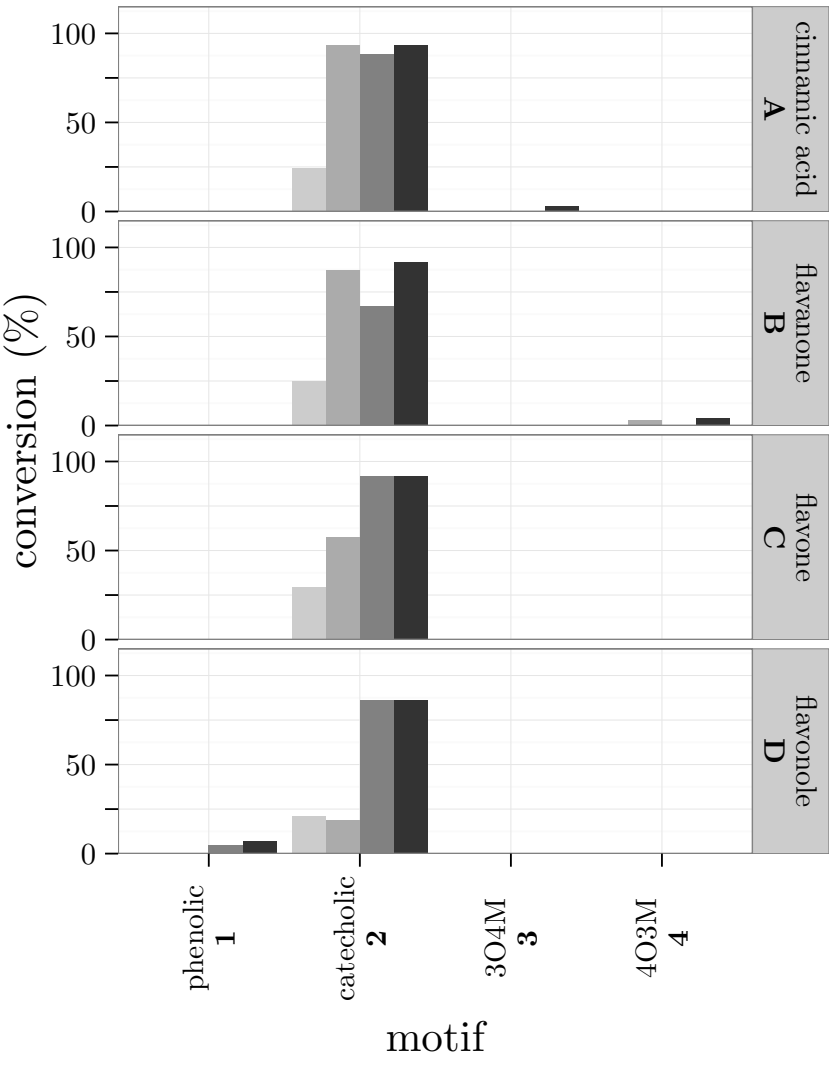


Figure 24: Same as above, but some data is omitted.

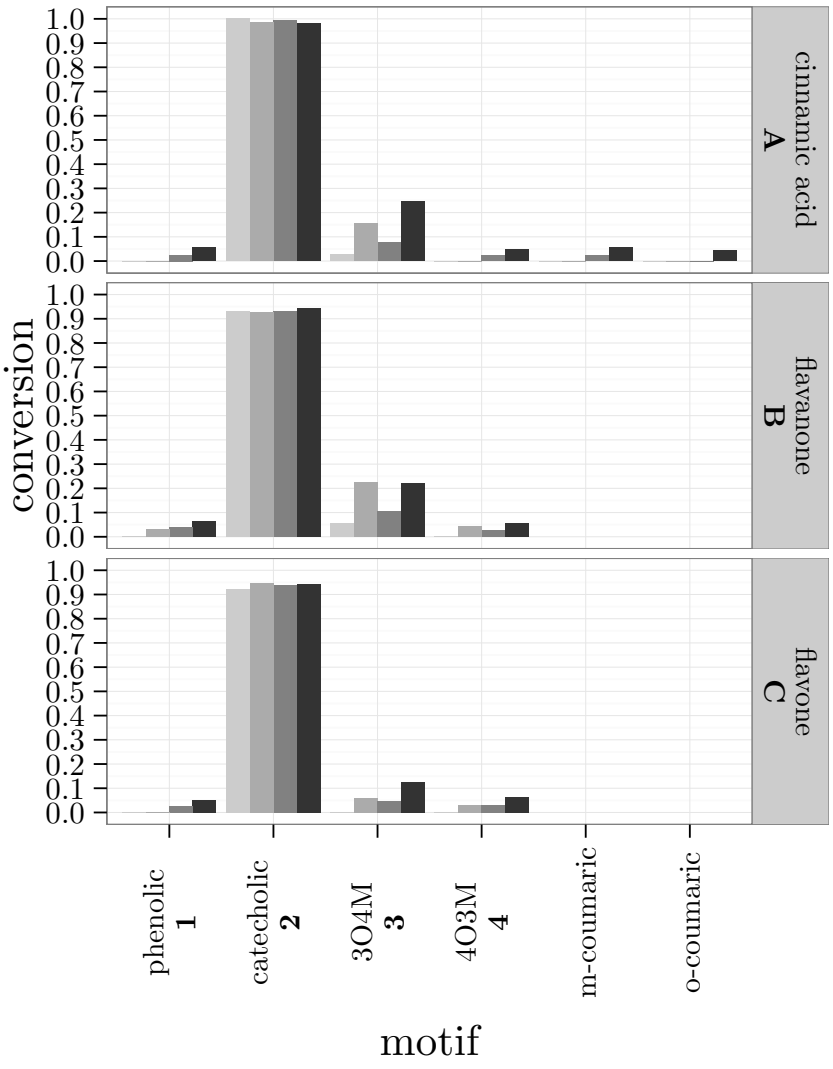
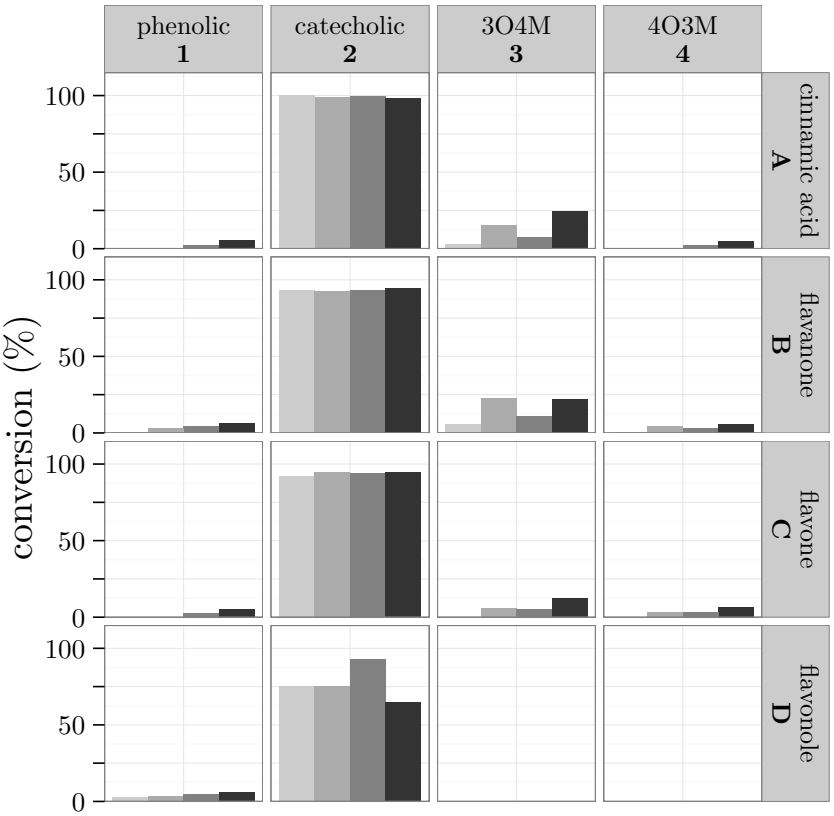
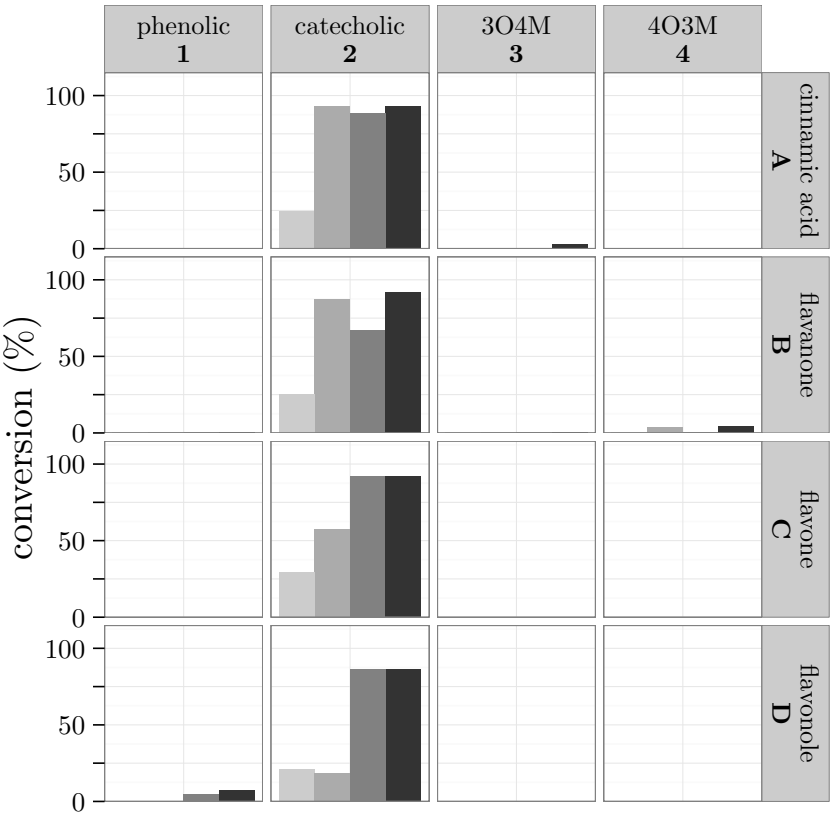


Figure 25: 4′-selective variant conversions. colors from light to dark: low/no, low/yes, high/no, high/yes


phenolic
1
catecholic
2
3O4M
3
4O3M
4
cinnamic acid
A
flavanone
B
flavone
C
flavonole
D
conversion (%)
100
50
0


phenolic
1
catecholic
2
3O4M
3
4O3M
4
cinnamic acid
A
flavanone
B
flavone
C
flavonole
D
conversion (%)
100
50
0

	substrate	variant	max	at.pH	at.Mg
1	1	Y51RN202W			
2	2	Y51RN202W	0.92	high	TRUE
3	3	Y51RN202W			
4	4	Y51RN202W	0.04	high	TRUE
5	5	Y51RN202W			
6	6	Y51RN202W	0.92	high	FALSE
7	7	Y51RN202W			
8	8	Y51RN202W			
9	9	Y51RN202W			
10	10	Y51RN202W			
11	11	Y51RN202W			
12	12	Y51RN202W	0.93	low	TRUE
13	14	Y51RN202W	0.03	high	TRUE
14	13	Y51RN202W			
15	15	Y51RN202W	0.07	high	TRUE
16	16	Y51RN202W	0.86	high	TRUE
17	17	Y51RN202W	1.02	low	TRUE