Electronic Supplementary Information (ESI)

# Insights of biomass recalcitrance in *Populus trichocarpa* natural variants for biomass conversion

Chang Geun Yoo,[a,b] Yongil Yang,[a] Yunqiao Pu,*[a] Xianzhi Meng,[c] Wellington Muchero,[a] Kelsey L. Yee,[a] Olivia A. Thompson,[a] Miguel Rodriguez Jr,[a] Garima Bali,[c] Nancy L. Engle,[a] Erika Lindquist,[d] Vasanth Singan,[d] Jeremy Schmutz,[d,e] Stephen P. DiFazio,[f] Timothy J. Tschaplinski,[a] Gerald A. Tuskan,[a] Jin-Gui Chen,[a] Brian Davison,[a] and Arthur J. Ragauskas*[a,b,g]

[a.] BioEnergy Science Center and Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA.
[b.] UT-ORNL Joint Institute for Biological Science, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA.
[c.] Renewable Bioproducts Institute, School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA.
[d.] Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598, USA.
[e.] HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA.
[f.] Department of Biology, West Virginia University, Morgantown, WV 26506, USA.
[g.] Department of Chemical and Biomolecular Engineering & Center for Renewable Carbon, Department of Forestry, Wildlife, and Fisheries, University of Tennessee, Knoxville, TN 37996, USA.
*Email: puy1@ornl.gov; aragausk@utk.edu

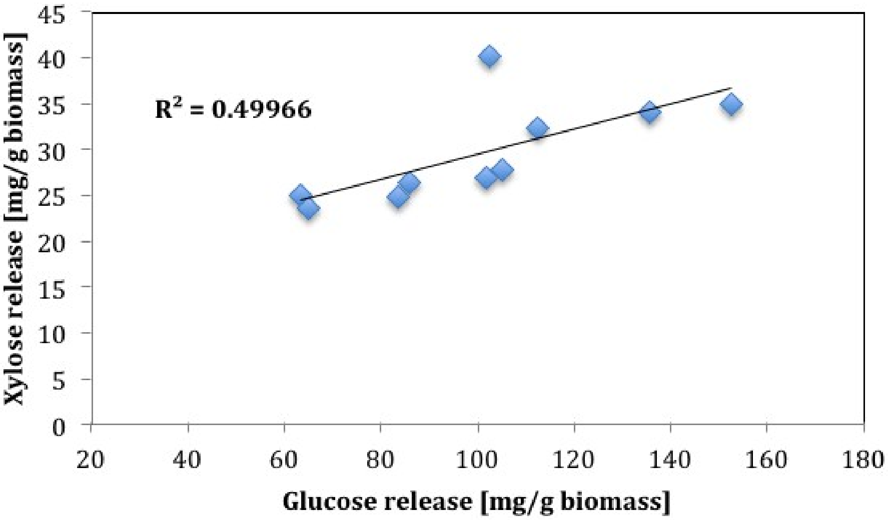

**Figure S1.** Correlation between glucose release and xylose release of *P. trichocarpa* natural variants ($R^2$=0.50, Pearson coefficient=0.705, *p*-value=0.023)

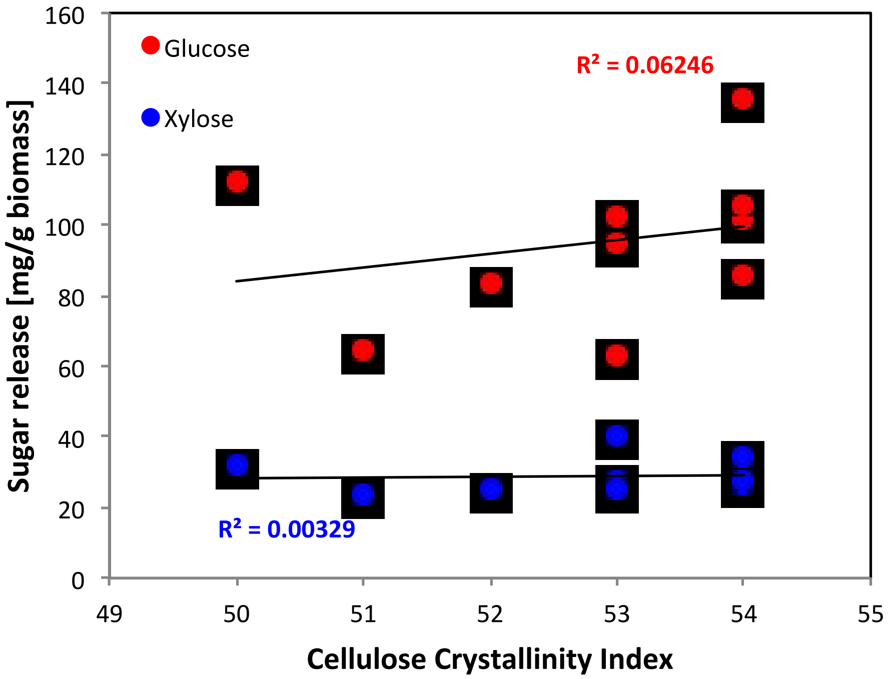

**Figure S2.** Correlation between sugar release and cellulose crystallinity of *P. trichocarpa* (CrI vs glucose release: $R^2$=0.06, Pearson coefficient=0.078, *p*-value=0.831; CrI vs xylose release: $R^2$=0.003, Pearson coefficient=-0.001, *p*-value=0.998)

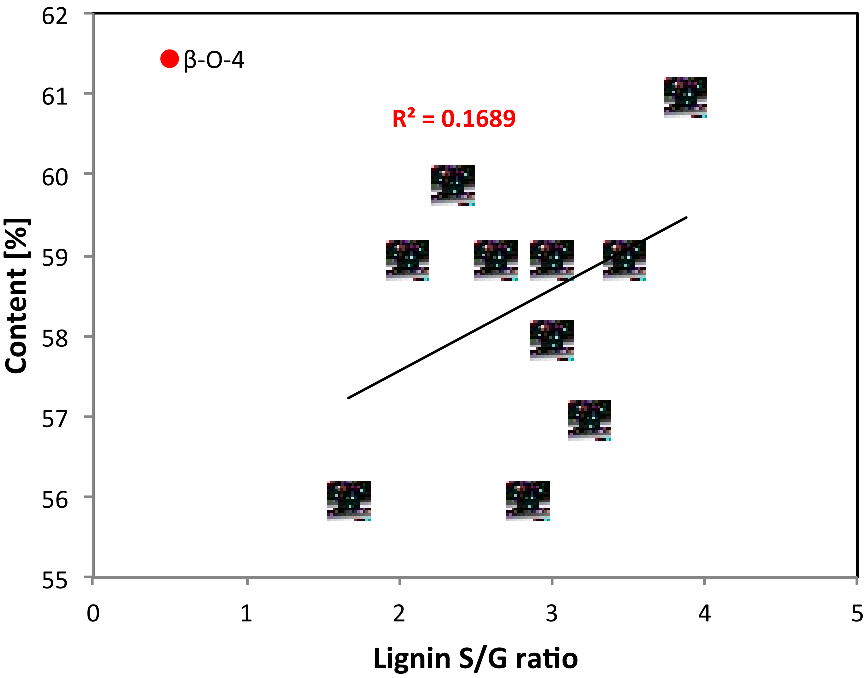

**Figure S3.** Correlation between lignin S/G ratio and β-*O*-4 linkage content ($R^2$=0.17, Pearson coefficient=0.411, *p*-value=0.238)

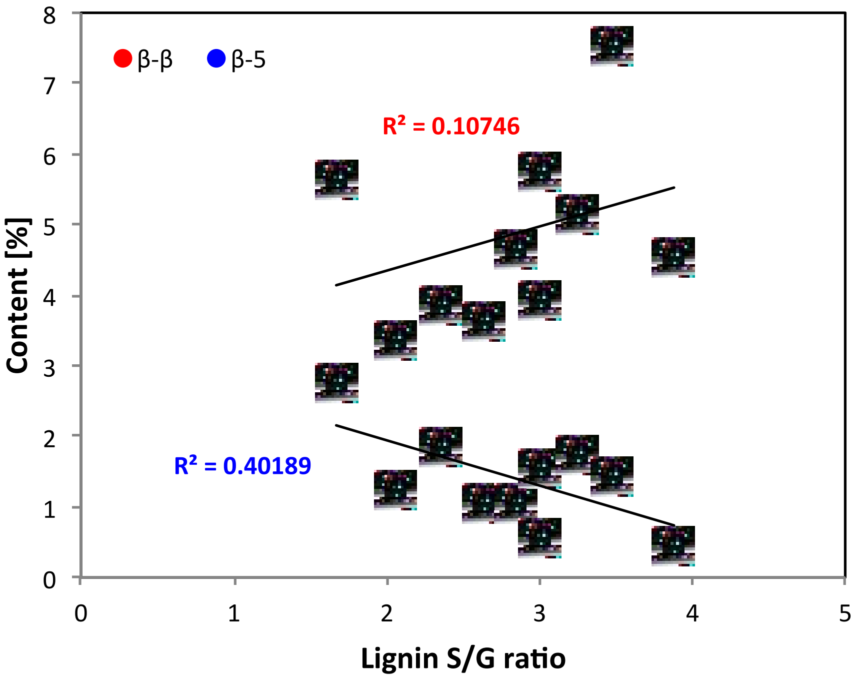

**Figure S4.** Correlation between lignin S/G ratio and β-5/β-β linkage content (S/G ratio vs β-5: $R^2$=0.40, Pearson coefficient=0.634, *p*-value=0.049; S/G ratio vs β-β: $R^2$=0.11, Pearson coefficient=0.328, *p*-value=0.355)

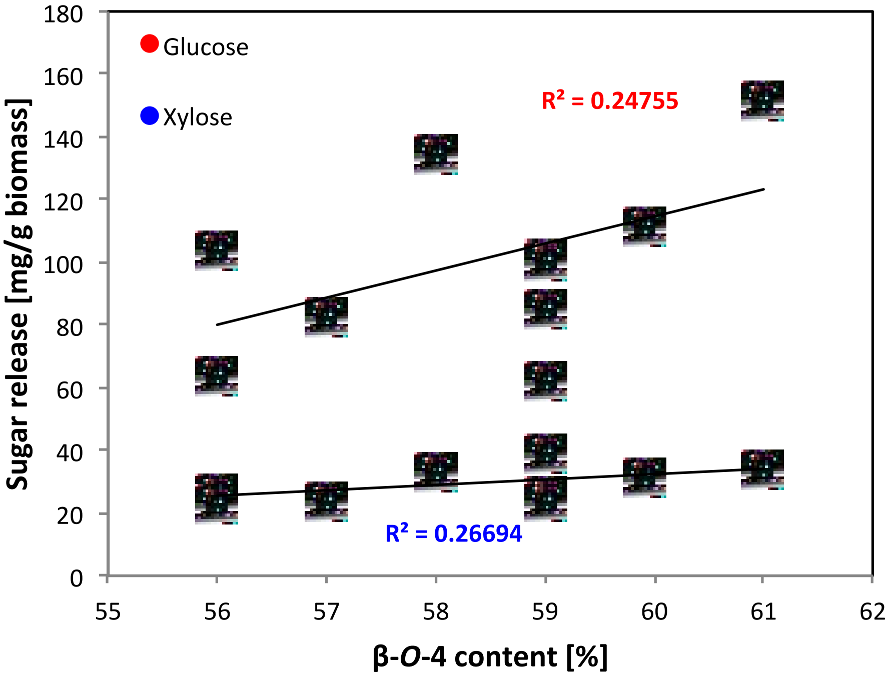

**Figure S5.** Correlation between β-*O*-4 linkage content and sugar release (β-*O*-4 vs glucose release: $R^2$=0.25, Pearson coefficient=0.498, *p*-value=0.143; β-*O*-4 vs xylose release: $R^2$=0.27, Pearson coefficient=0.517, *p*-value=0.126)

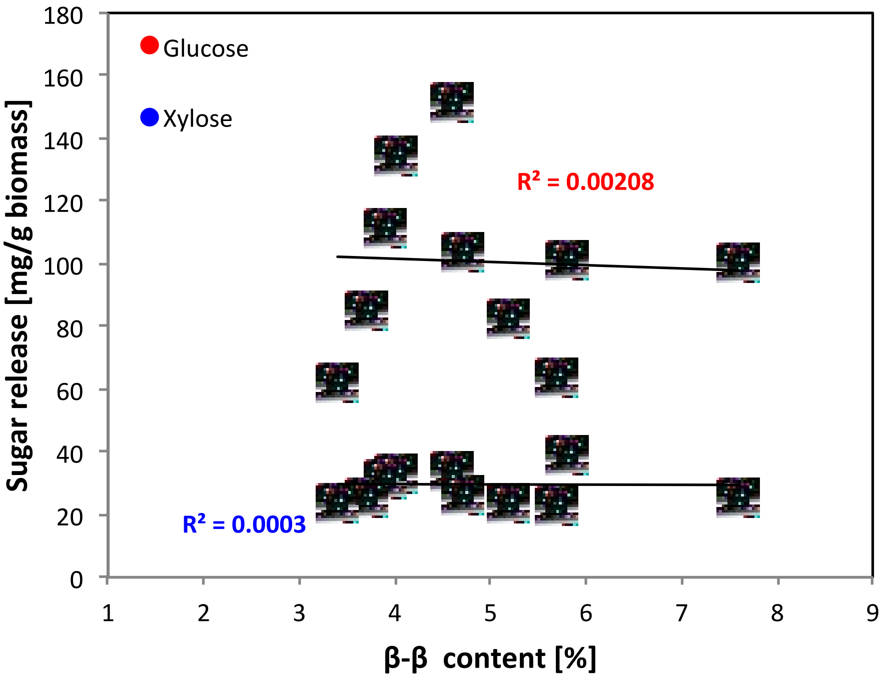

**Figure S6.** Correlation between β-β linkage content and sugar release (β-β vs glucose release: $R^2$=0.002, Pearson coefficient=-0.045, *p*-value=0.901; β-β vs xylose release: $R^2$=0.0003, Pearson coefficient=-0.017, *p*-value=0.963)

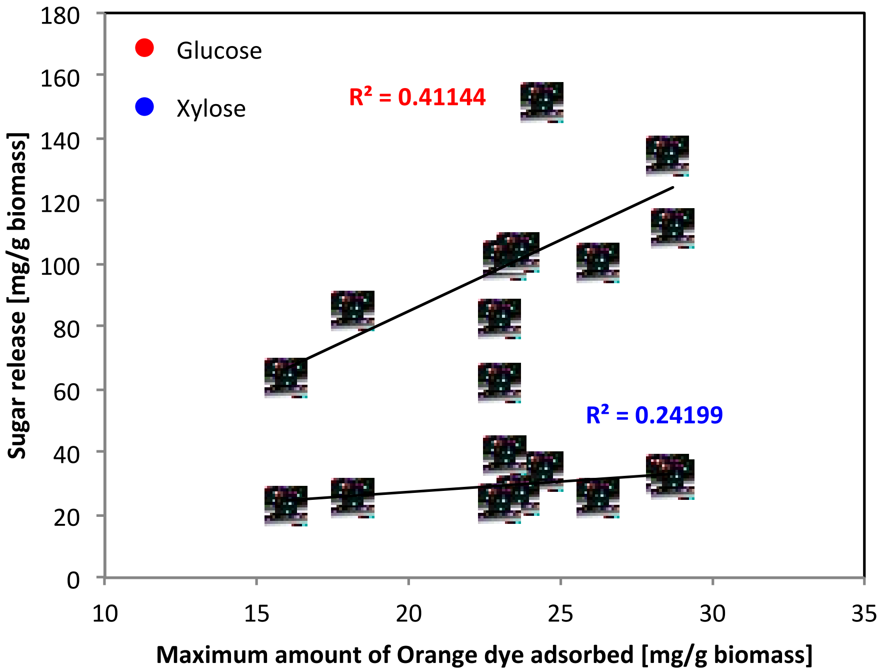

**Figure S7.** Correlation between maximum amount of orange dye adsorption and sugar release (Orange dye vs glucose release: $R^2$=0.41, Pearson coefficient=0.641, *p*-value=0.046; Orange dye vs xylose release: $R^2$=0.24, Pearson coefficient=0.493, *p*-value=0.148)

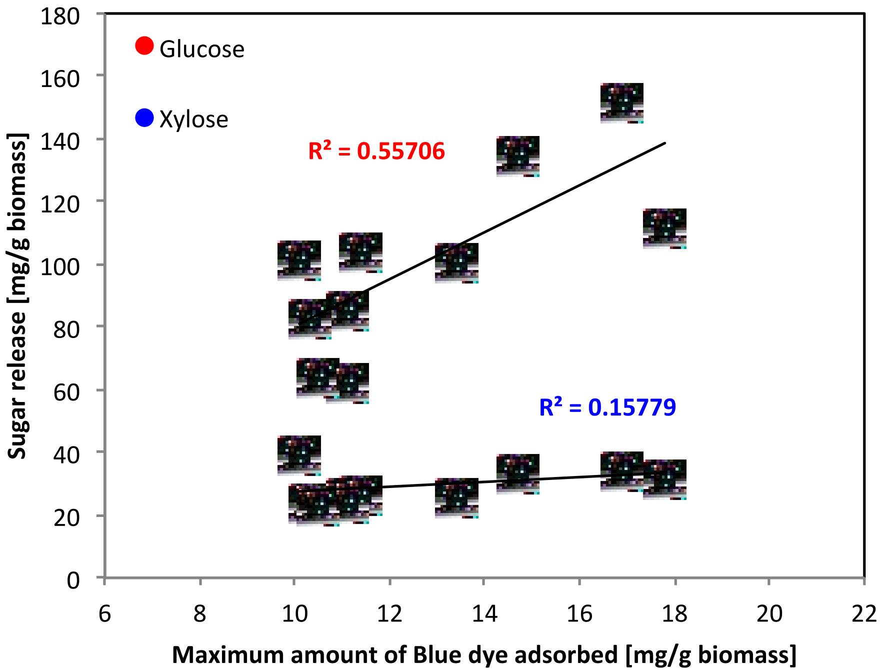

**Figure S8.** Correlation between maximum amount of blue dye adsorption and sugar release (Blue dye vs glucose release: $R^2$=0.56, Pearson coefficient=0.747, *p*-value=0.013; Blue dye dye vs xylose release: $R^2$=0.16, Pearson coefficient=0.397, *p*-value=0.256)

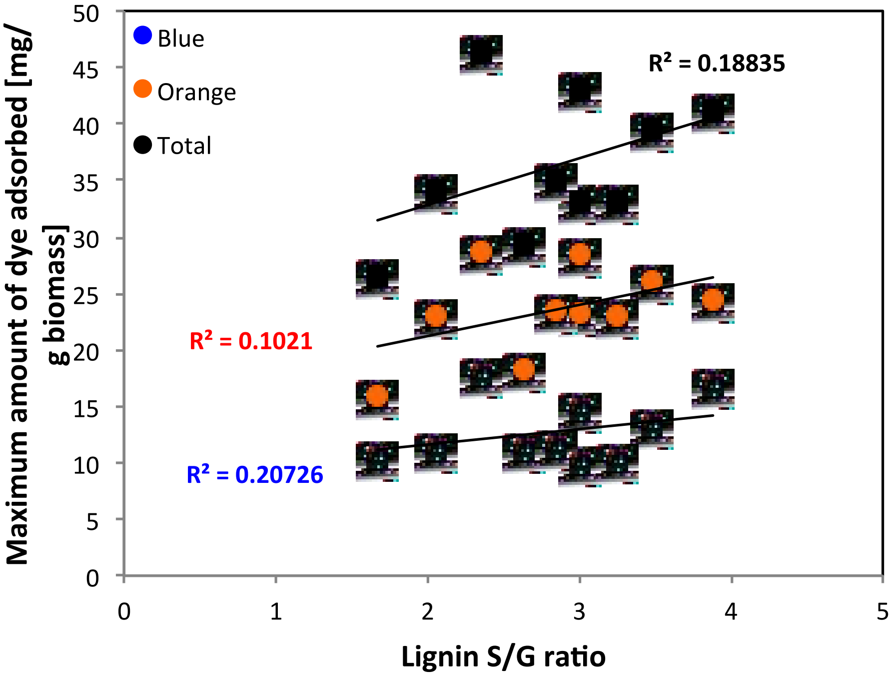

**Figure S9.** Correlation between cellulose accessibility and lignin S/G ratio

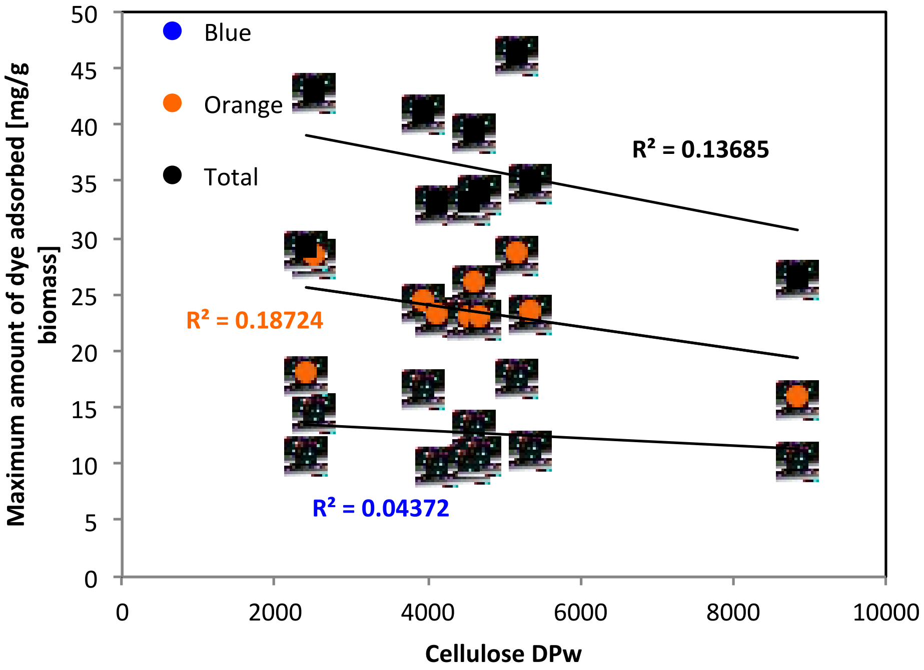

**Figure S10.** Correlation between cellulose accessibility and cellulose $DP_w$

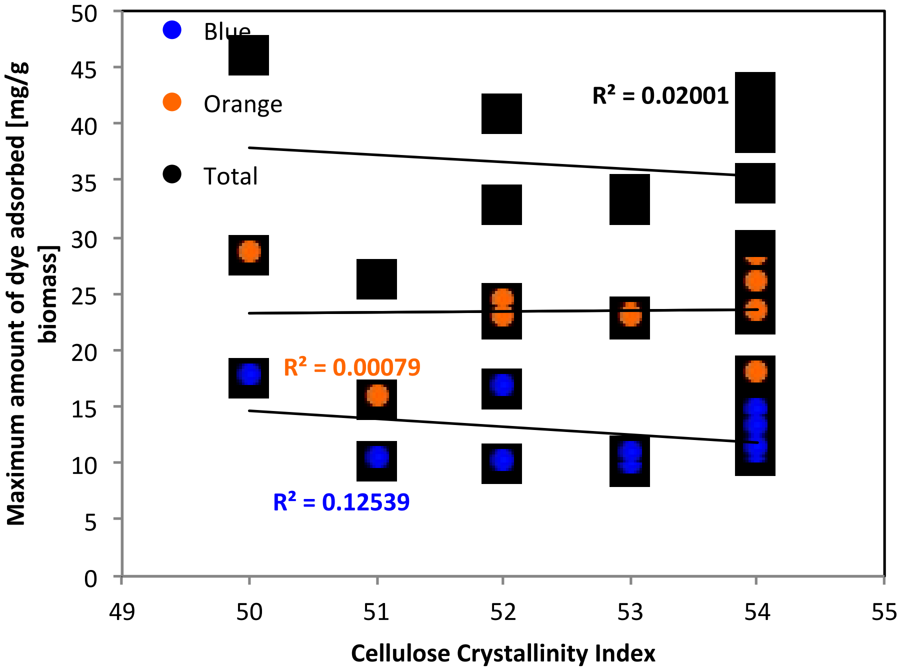

**Figure S11.** Correlation between cellulose accessibility and cellulose crystallinity index

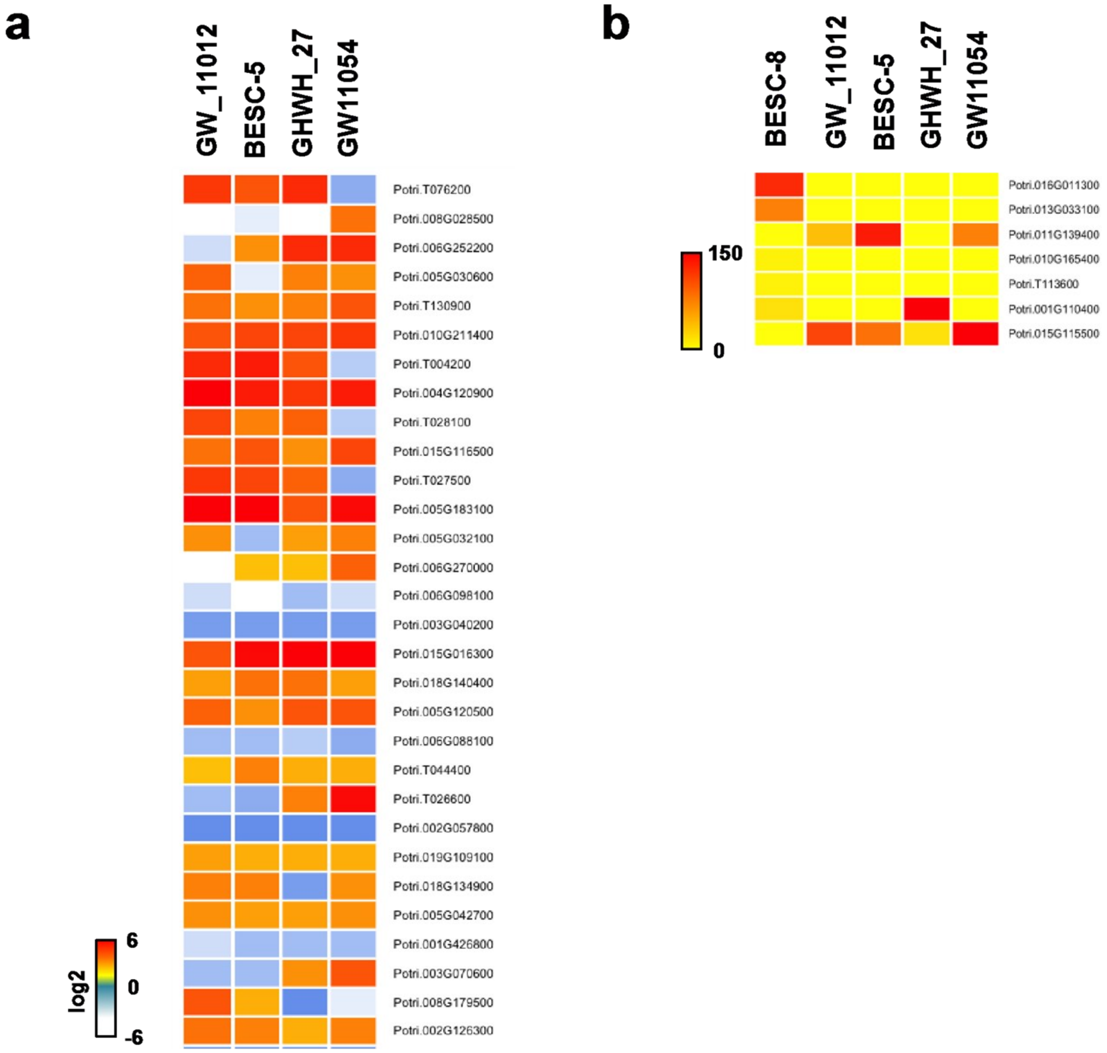

**Figure S12.** Differentially expressed genes between *P. trichocarpa* natural variant genotypes

## Table S1. The list of correlation efficient between physicochemical property analysis and differentially expressed genes over natural variant.

| geneID | 11012vs8 | 27vs8 | 8vs8 | 5 vs8 | 11054vs8 | lignin content | S/G ratio | celluloseDPw | glucoserelease | xyloserelease | ligninMw | hemi | description | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Potri.008G179500 | 3.81975293 | -1.332104 | 1 | 1.84359984 | -5.691667 | -0.911323751 | 0.85581463 | -0.674709388 | 0.952825203 | 0.572087537 | 0.54137444 | 0.609828487 | multidrug resistance-associated protein 3 | |
| Potri.T028100 | 4.33235249 | 3.58601314 | 1 | 2.87452429 | -4.1472149 | -0.766462965 | 0.61796706 | -0.452174717 | 0.854398316 | 0.28022806 | 0.802255359 | 0.744175934 | NB-ARC domain-containing disease resistance protein | |
| Potri.T027500 | 4.52750026 | 3.62077982 | 1 | 4.39221715 | -2.9935358 | -0.727351302 | 0.69230093 | -0.391721398 | 0.816870684 | 0.157105432 | 0.810267066 | 0.696825896 | NB-ARC domain-containing disease resistance protein | |
| Potri.T004200 | 4.98984575 | 3.86690171 | 1 | 5.21454249 | -4.3588461 | -0.720085421 | 0.72477041 | -0.37508968 | 0.813054971 | 0.178186203 | 0.823054585 | 0.719100278 | NB-ARC domain-containing disease resistance protein | |
| Potri.T076200 | 4.54896256 | 5.01230443 | 1 | 3.83069045 | -2.8470997 | -0.66083187 | 0.53135559 | -0.332799942 | 0.758156442 | 0.107269627 | 0.839155931 | 0.715284008 | Zinc-binding dehydrogenase family protein | |
| Potri.019G109100 | 2.18179376 | 1.65543109 | 1 | 1.58330057 | 1.64916644 | -0.469442202 | -0.0843557 | -0.556489882 | 0.398253632 | -0.388506516 | -0.145771316 | -0.394934515 | | 0 |
| Potri.019G000200 | 6.9398581 | -3.2817515 | 1 | -3.2817515 | 2.90357465 | -0.491085196 | -0.1351334 | -0.810679683 | 0.344041264 | 0.371190686 | -0.697867321 | -0.549661252 | | 0 |
| Potri.018G134900 | 2.92292587 | -2.0278906 | 1 | 2.75074805 | 2.35771369 | -0.280975196 | 0.39781009 | -0.42505571 | 0.17741663 | 0.068813487 | -0.507645112 | -0.498337891 | Protein kinase superfamily protein | |
| Potri.001G110400 | -3.6706253 | 4.08316387 | 1 | -3.6706253 | -5.7477199 | -0.005124478 | -0.0585862 | 0.220713343 | 0.127334409 | 0.272850065 | 0.683722038 | 0.77105134 | sugar transporter 1 | |
| Potri.004G120900 | 6.25767115 | 4.68146046 | 1 | 5.36246632 | 5.5293851 | -0.141446235 | -0.1387976 | -0.213398899 | 0.08134999 | -0.702414787 | -0.178889066 | -0.52218854 | Domain of unknown function (DUF3598) | |
| Potri.T113600 | -3.2300527 | -3.2300527 | 1 | -3.2300527 | -3.2300527 | -0.031805917 | 0.22617267 | -0.013787495 | 0.059102421 | 0.83885322 | 0.039968038 | 0.427337218 | P-loop containing nucleoside triphosphate hydrolases superfamily protein | |
| Potri.010G165400 | -2.8450347 | -2.8450347 | 1 | -2.8450347 | -2.8450347 | -0.031805917 | 0.22617267 | -0.013787495 | 0.059102421 | 0.83885322 | 0.039968038 | 0.427337218 | EXS (ERD1/XPR1/SYG1) family protein | |
| Potri.005G030600 | 3.63711338 | 2.73270779 | 1 | -5.7401287 | 2.46318183 | -0.128721495 | -0.6829662 | -0.406059754 | 0.045988182 | 0.19851534 | -0.43781429 | -0.315288475 | Leucine-rich receptor-like protein kinase family protein | |
| Potri.005G183100 | 5.94793944 | 4.02036985 | 1 | 5.95382289 | 5.5731498 | -0.100365885 | -0.0157232 | -0.155389823 | 0.041494432 | -0.70843889 | -0.193364021 | -0.535610771 | | 0 |
| Potri.002G126300 | 3.05730255 | 1.76455318 | 1 | 2.86906579 | 2.94803383 | -0.130889708 | 0.00053713 | -0.256134871 | 0.041394834 | -0.589148703 | -0.37900703 | -0.657502469 | alpha-amylase-like | |
| Potri.006G098100 | -5.1791433 | -3.3565611 | 1 | -6.9759709 | -5.3616988 | -0.000391913 | -0.0617737 | -0.004777885 | 0.034691228 | 0.769021699 | 0.087224019 | 0.451470499 | plasma membrane intrinsic protein 1;4 | |
| Potri.005G042700 | 2.5077552 | 2.06467857 | 1 | 2.25070522 | 2.37699846 | -0.098473086 | -0.1812805 | -0.188509669 | 0.033608509 | -0.721532245 | -0.222135775 | -0.563694496 | DCD (Development and Cell Death) domain protein | |
| Potri.013G033100 | -5.7264785 | -3.5664764 | 1 | -6.5203475 | -5.7264785 | 0.0279542 | 0.01727431 | 0.063600707 | 0.020839789 | 0.760503696 | 0.15863678 | 0.516433227 | Esterase/lipase/thioesterase family protein | |
| Potri.006G088100 | -3.3532627 | -4.0814108 | 1 | -3.6789301 | -3.0491148 | -0.017484954 | 0.18001595 | -0.067701697 | 0.015777669 | 0.830896031 | -0.134914446 | 0.26356908 | Pathogenesis-related thaumatin superfamily protein | |
| Potri.003G040200 | -1.9542864 | -1.5711539 | 1 | -1.7590178 | -1.6978483 | 0.057700184 | 0.16737847 | 0.085660369 | -0.020410432 | 0.783389195 | 0.068873533 | 0.443831996 | Metallo-hydrolase/oxidoreductase superfamily protein | |
| Potri.002G057800 | -1.4315601 | -1.1900929 | 1 | -1.0884879 | -1.180846 | 0.065873897 | 0.22570138 | 0.105158474 | -0.027317468 | 0.770630941 | 0.072680488 | 0.440902498 | vacuoleless1 (VCL1) | |
| Potri.015G115500 | 6.24558649 | 3.76562407 | no_detection | 5.84613807 | 6.71476219 | -0.06146801 | 0.09754135 | -0.329518352 | -0.071504973 | 0.136869268 | -0.771052611 | -0.77156665 | | 0 |
| Potri.005G032100 | 2.5692233 | 2.20010334 | 1 | -3.5287406 | 2.72990377 | -0.007216905 | -0.7664942 | -0.317114883 | -0.083793786 | 0.117028276 | -0.526397268 | -0.415319303 | peptidoglycan-binding LysM domain-containing protein | |
| Potri.018G140400 | 1.97372042 | 3.24724233 | 1 | 3.02719138 | 2.08209585 | 0.186054992 | -0.0434646 | 0.400795468 | -0.107353309 | -0.801804213 | 0.504095817 | 0.118677824 | UDP-Glycosyltransferase superfamily protein | |
| Potri.011G139400 | 4.5953431 | 2.10415823 | no_detection | 6.55673881 | 5.61751557 | 0.066280752 | 0.44709891 | 0.009504909 | -0.12242907 | -0.116205041 | -0.365167219 | -0.413244953 | | 0 |
| Potri.016G011300 | -8.7709759 | -6.4622083 | 1 | -6.4622083 | -6.4622083 | 0.210838285 | 0.14242706 | 0.254409011 | -0.163098092 | 0.666475043 | 0.079976573 | 0.423446742 | Mo25 family protein | |
| Potri.T044400 | 1.49551303 | 1.5946225 | 1 | 2.90435474 | 1.86769276 | 0.201418731 | 0.34172013 | 0.367554842 | -0.166111171 | -0.687777168 | 0.211013864 | -0.109582142 | | 0 |
| Potri.001G426800 | -4.8802268 | -3.397298 | 1 | -3.397298 | -3.397298 | 0.22945999 | 0.13532616 | 0.274924816 | -0.180244714 | 0.650689924 | 0.082905449 | 0.421811631 | | 0 |
| Potri.005G120500 | 3.55078969 | 3.80193842 | 1 | 2.53914438 | 3.96274838 | 0.141050247 | -0.6044791 | -0.014165621 | -0.200351809 | -0.76166117 | -0.277535341 | -0.584184961 | Glycosyl hydrolase superfamily protein | |
| Potri.010G211400 | 3.78108519 | 4.38468666 | 1 | 4.1014219 | 4.55329346 | 0.225678514 | -0.3704118 | 0.18356334 | -0.250255671 | -0.921462901 | -0.094207937 | -0.482455771 | sulfate transporter 4.1 | |
| Potri.T130900 | 3.19295559 | 2.87591547 | 1 | 2.37636811 | 3.84034765 | 0.180421063 | -0.5851295 | -0.045774333 | -0.272670052 | -0.752662975 | -0.49175594 | -0.774526172 | Tetratricopeptide repeat (TPR)-like superfamily protein | |
| Potri.015G116500 | 3.07073495 | 2.54089415 | 1 | 3.93845341 | 4.12540016 | 0.247891767 | -0.1529626 | 0.158226657 | -0.305310205 | -0.84198645 | -0.317323126 | -0.655856642 | Protein of unknown function (DUF594) | |
| Potri.015G016300 | 3.94858441 | 5.93569236 | 1 | 5.80039967 | 6.73208928 | 0.469855145 | -0.4314128 | 0.429409476 | -0.482472842 | -0.991445961 | -0.121332077 | -0.500536001 | OBP3-responsive gene 1 | |
| Potri.008G028500 | -6.3948821 | -6.2704362 | 1 | -5.9621775 | 3.02177675 | 0.608060531 | -0.4823248 | 0.311235419 | -0.686037406 | 0.131121441 | -0.739134858 | -0.517454024 | Phosphorylase superfamily protein | |
| Potri.003G070600 | -3.6250868 | 2.31419501 | 1 | -3.5294423 | 3.76779813 | 0.795613457 | -0.8723115 | 0.592206335 | -0.810640103 | -0.229322834 | -0.384620206 | -0.317125513 | Chaperone DnaJ-domain superfamily protein | |
| Potri.006G252200 | -5.3358116 | 5.11364703 | 1 | 2.60068421 | 4.95144963 | 0.94244724 | -0.5010894 | 0.977094937 | -0.878940417 | -0.598915997 | 0.056596366 | -0.068631015 | plant natriuretic peptide A | |
| Potri.T026600 | -3.7646697 | 2.99665736 | 1 | -2.9271235 | 5.8312303 | 0.857266971 | -0.8908425 | 0.637040187 | -0.881910489 | -0.35838261 | -0.448929102 | -0.428673763 | LRR and NB-ARC domains-containing disease resistance protein | |
| Potri.006G270000 | -7.4038908 | 1.41114792 | 1 | 1.52417933 | 3.54704961 | 0.953609402 | -0.3642964 | 0.969350926 | -0.90072064 | -0.431172004 | -0.041208476 | -0.081990829 | Disease resistance protein (TIR-NBS-LRR class) family | |
| lignincontent | 17.83 | 20.63 | 20.03 | 20.17 | 21.91 | 1 | -0.6011638 | 0.906876104 | -0.986250862 | -0.557134148 | -0.266108747 | -0.338388836 | | |
| S/G ratio | 3 | 2.38 | 3.01 | 3.47 | 2.05 | -0.601163843 | 1 | -0.312190106 | 0.664198076 | 0.42017513 | 0.546662628 | 0.587448213 | | |
| celluloseDPw | 2505 | 4721 | 4104 | 4624 | 4682 | 0.906876104 | -0.3121901 | 1 | -0.827303834 | -0.540679229 | 0.152367846 | 0.037676617 | | |
| glucose release | 135.5 | 95 | 102.3 | 101.9 | 63.2 | -0.986250862 | 0.66419808 | -0.827303834 | 1 | 0.55590743 | 0.420264562 | 0.482370448 | | |
| xylose release | 34.1 | 27 | 40.2 | 27 | 25 | -0.557134148 | 0.42017513 | -0.540679229 | 0.55590743 | 1 | 0.07816012 | 0.450623395 | | |
| ligninMw | 10900 | 11600 | 11100 | 11500 | 10200 | -0.266108747 | 0.54666263 | 0.152367846 | 0.420264562 | 0.07816012 | 1 | 0.914448692 | | |
| hemicelluloseDPw | 373 | 405 | 407 | 396 | 328 | -0.338388836 | 0.58744821 | 0.037676617 | 0.482370448 | 0.450623395 | 0.914448692 | 1 | | |