

Integrating genetic variation with metabolite abundance and gene expression in *Populus tremula*

Niklas Mähler, Kathryn M. Robinson, Torgeir R. Hvidsten, Nathaniel R. Street

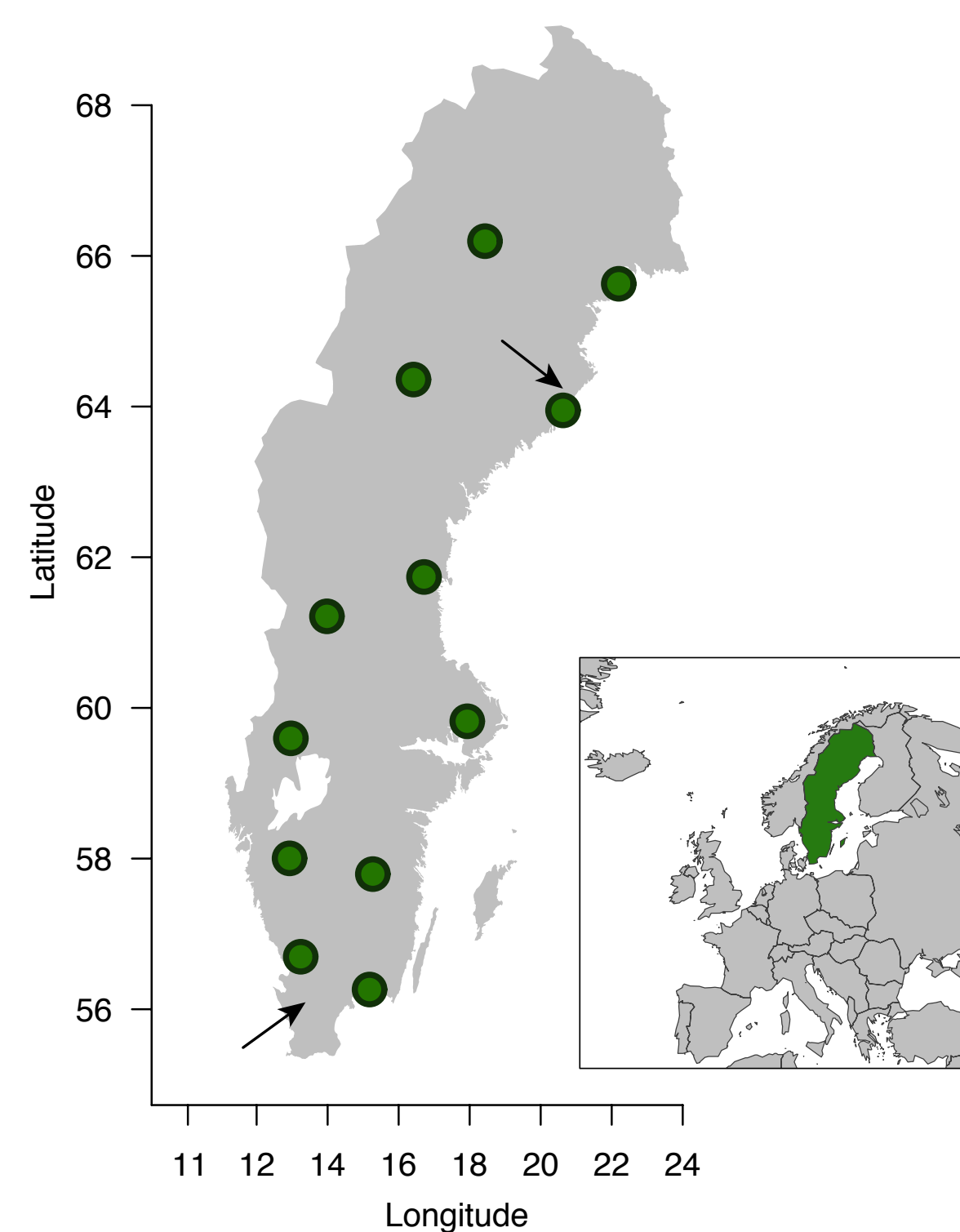


Salicinoid-like phenolic glycosides, or salicinoids, are specialised metabolites that can constitute 20% of leaf dry weight in *Populus tremula*. They are biologically active compounds that are associated with herbivore resistance, amongst other things. The biogenesis pathways for these compounds are largely unknown. We aim to integrate metabolite abundance with gene expression and genetic variation in order to suggest candidate genes and possible mechanisms behind the biosynthesis of these compounds.

The Swedish Aspen Collection

Aspen trees (*Populus tremula*) were collected from all across Sweden to form the Swedish Aspen Collection (SwAsp [1]). A total of 116 individuals were collected and set up in two common gardens, one in northern Sweden and one in southern Sweden (figure 1). Data in the form of DNA-sequencing data, RNA-sequencing data, and salicinoid abundance (LCMS) were collected. After quality filtering, 80 distinct genotypes could be used for downstream analysis. The abundance of 25 different salicinoids was estimated, and it has been shown previously that these metabolites does not show any geographical structure in this population [2]. Genetic variation in this population has been described previously [3].

Figure 1. Map of the Swedish Aspen Collection. Aspens were collected from 12 locations in Sweden (green dots), consisting of a total of 116 individuals. Two common gardens were set up, one in the north and one in the south (marked with arrows).



Correlate all the things

We associated genetic variation in the form of SNPs with gene expression using the R-package Matrix eQTL [4,5], and we performed GWAS using GEMMA [6]. We have previously tried performing GWAS on higher order phenotypes, such as leaf shape, which practically yielded no significant associations. However, for both mRNA levels and metabolite abundance (example in figure 2) we find significant associations, indicating that these have a simpler genetic architecture than leaf shape. It might also indicate that these molecular phenotypes are under positive selection while the more complex traits are under neutral selection. A total of 6241 genes have SNPs associated with their expression level and 15 out of 25 salicinoids have significant peaks in the GWAS.

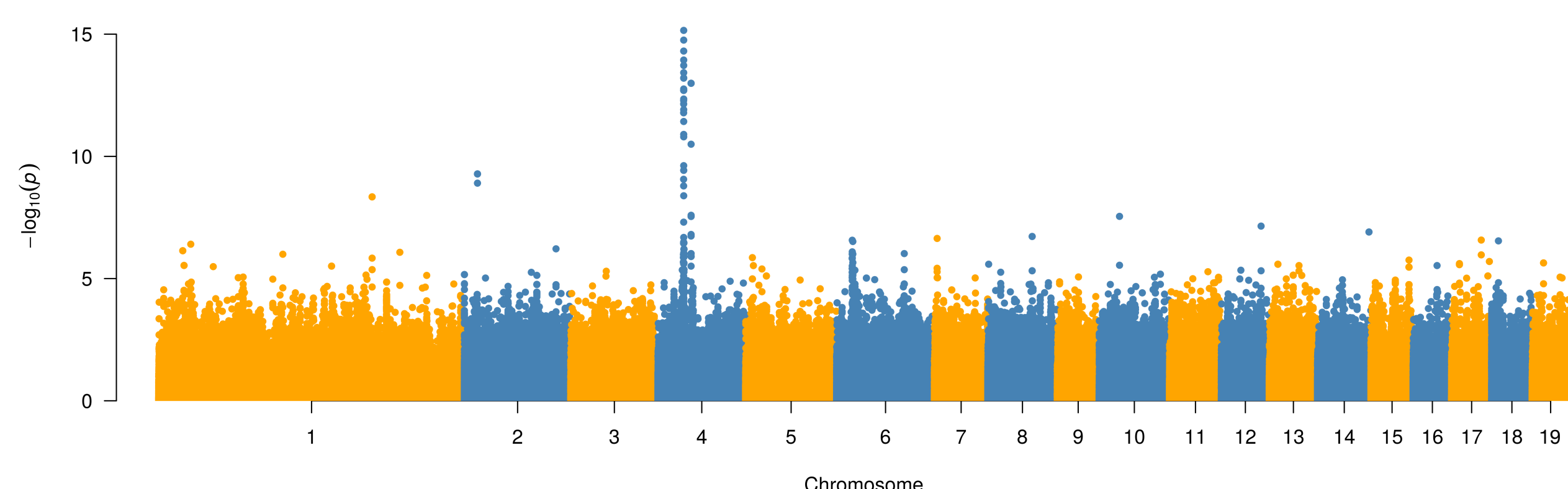


Figure 2. Manhattan plot for the GWAS of the salicinoid cinnamoyl salicortin.

Putting everything together

In figure 3, metabolites are visualised together with genes. The set of genes was chosen by taking either the genes that are co-located with significant GWAS SNPs, or genes whose expression is also associated with significant GWAS SNPs. Highly connected genes in this example could be interesting targets for functional studies using e.g. CRISPR/Cas9.

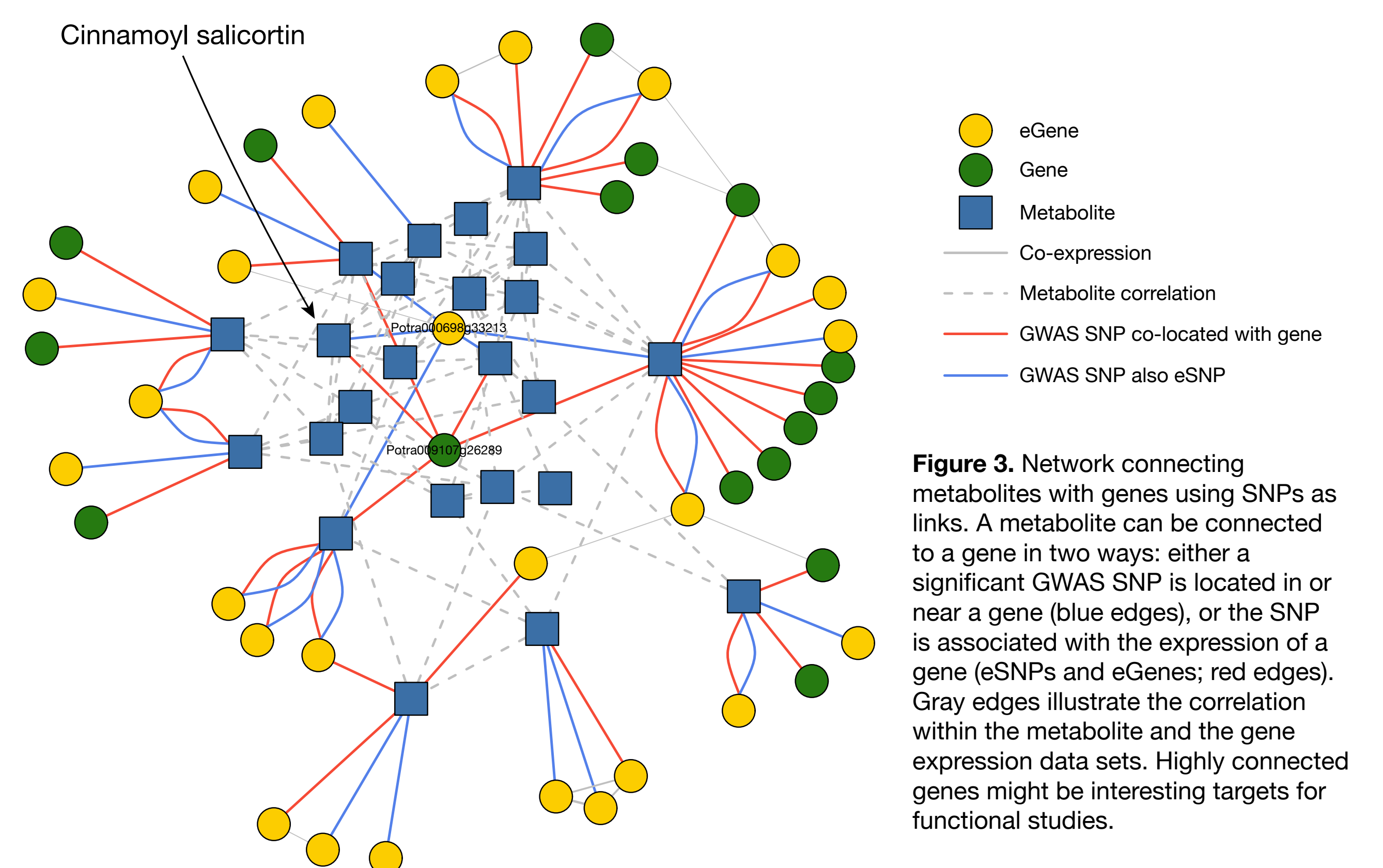
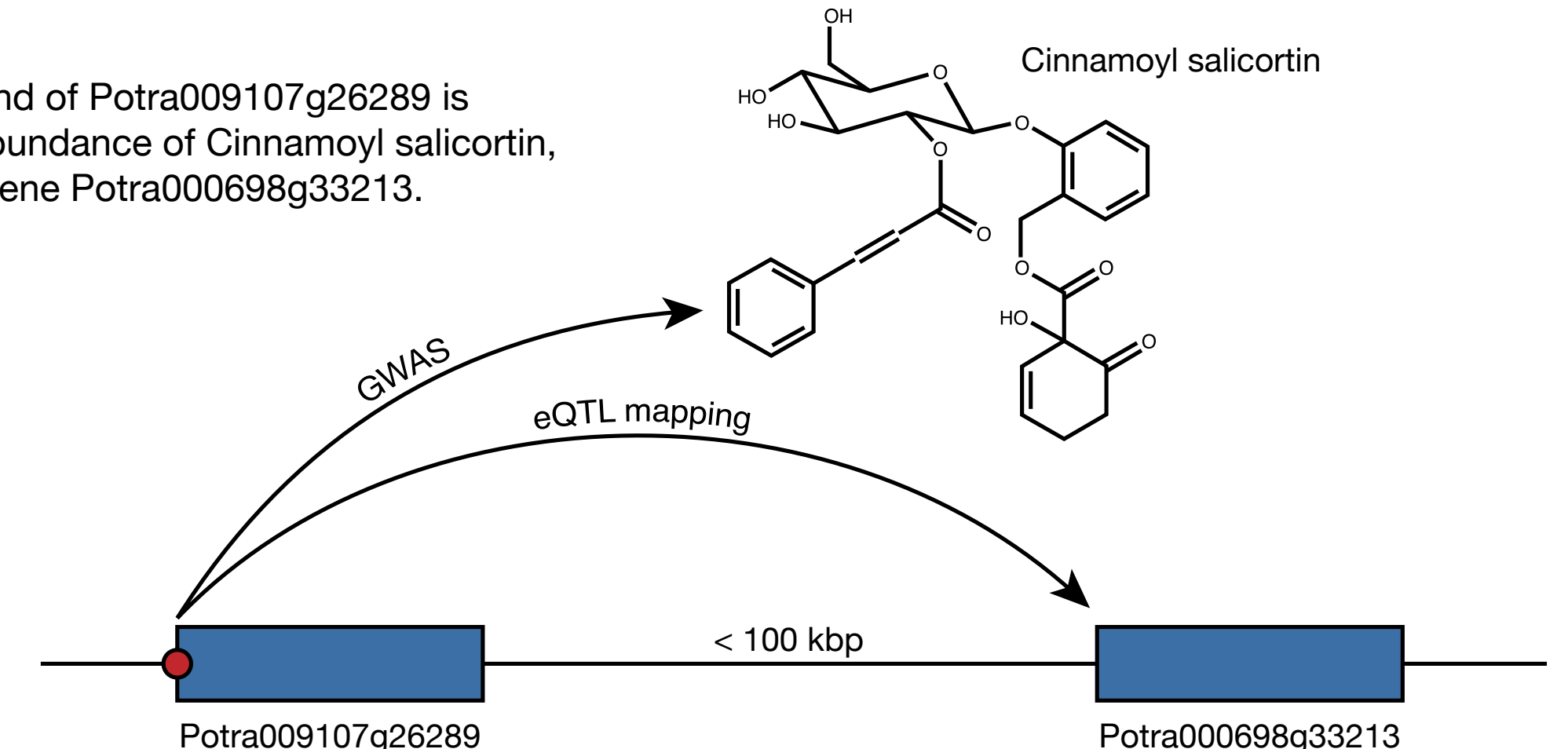


Figure 3. Network connecting metabolites with genes using SNPs as links. A metabolite can be connected to a gene in two ways: either a significant GWAS SNP is located in or near a gene (blue edges), or the SNP is associated with the expression of a gene (eSNPs and eGenes; red edges). Gray edges illustrate the correlation within the metabolite and the gene expression data sets. Highly connected genes might be interesting targets for functional studies.

A more detailed example

The salicinoid cinnamoyl salicortin (marked in figure 3) is associated with the genes Potra009107g26289 and Potra000698g33213 from the GWAS and the eQTL mapping, respectively (figure 4). Potra009107g26289 is annotated as a beta-carotene hydroxylase and Potra000698g33213 as an acyl transferase protein.

Figure 4. A SNP in the 5' end of Potra009107g26289 is associated with both the abundance of Cinnamoyl salicortin, and the expression of the gene Potra000698g33213.



Making things public

We are working on a web interface for exploring these kinds of data. Instead of having something static, like this poster, we want to make it possible for anyone to explore this. This will eventually become a part of PlantGenie.org [7].

References

1. Luquez et al. Tree Genetics & Genomes 2008. DOI: 10.1007/s11295-007-0108-y
2. Keefover-Ring et al. PLOS ONE 2014. DOI: 10.1371/journal.pone.0107189
3. Wang et al. Genetics 2016. DOI: 10.1534/genetics.115.183152
4. Mähler et al. PLOS Genetics 2017. DOI: 10.1371/journal.pgen.1006402
5. Shabalin. Bioinformatics 2012. DOI: 10.1093/bioinformatics/bts163
6. Zhou & Stephens. Nature Genetics 2012. DOI: 10.1038/ng.2310
7. Sundell et al. New Phytologist 2015. DOI: 10.1111/nph.13557