Using Hierarchical Clustering Of Metabolomic Data To Relate Clusters To Metabolite Pathways

Susan Verrall, Jim McNicol, Tom Shepherd, Gary Dobson, D.Wynne Griffiths, Gavin Ramsay, Howard Davies, and Derek Stewart
SCRI, Errol Road, Invergowrie, DD2 5DA

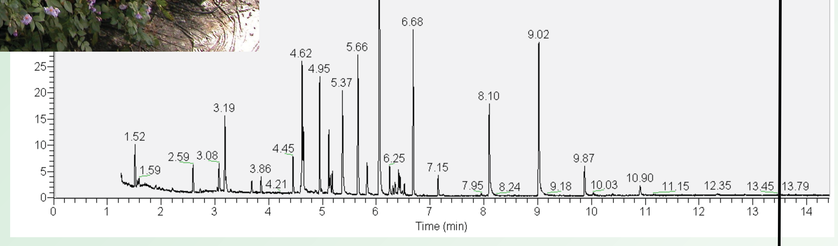
Introduction

Metabolic compositional profiles, consisting of hundreds of compound intensities, are increasingly used at SCRI to characterise samples of different genotypic and environmental backgrounds. Using hierarchical clustering we have grouped significant metabolites, and in turn these groups of compounds have been related to metabolite pathways.

Potato material

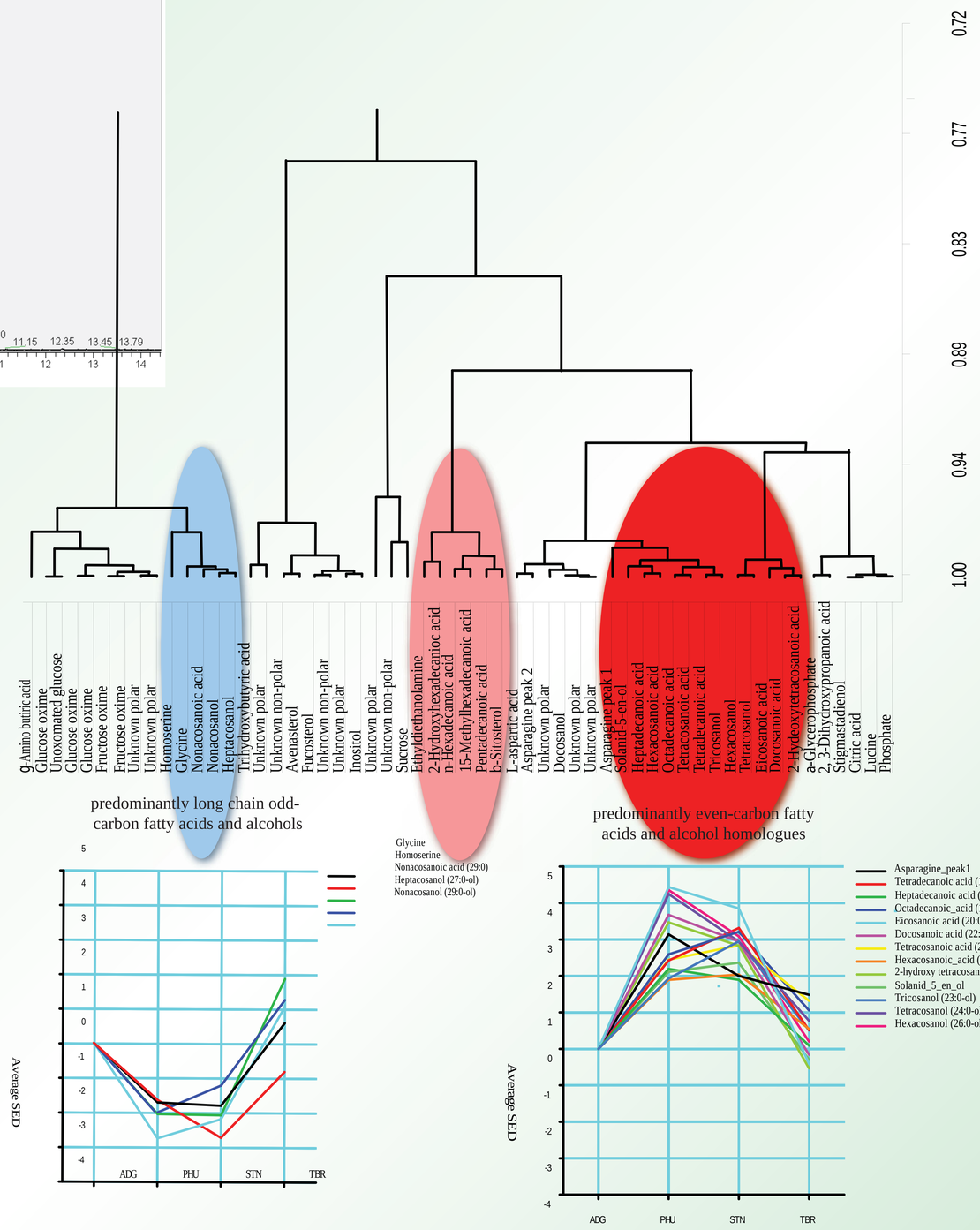
Plants representing the four main cultivated groups of potato within the broad definition of *Solanum tuberosum*, Andigena (ADG), Phureja (PHU), Stenotomum (STN) and Chilean Tuberosum (TBR), were grown from true seed in a glasshouse, with 78, 43, 24 and 11 accessions, respectively.

GC-MS analysis was performed, and profiles consisting of 78 GC-MS polar and 52 non-polar compounds were used in the analysis.



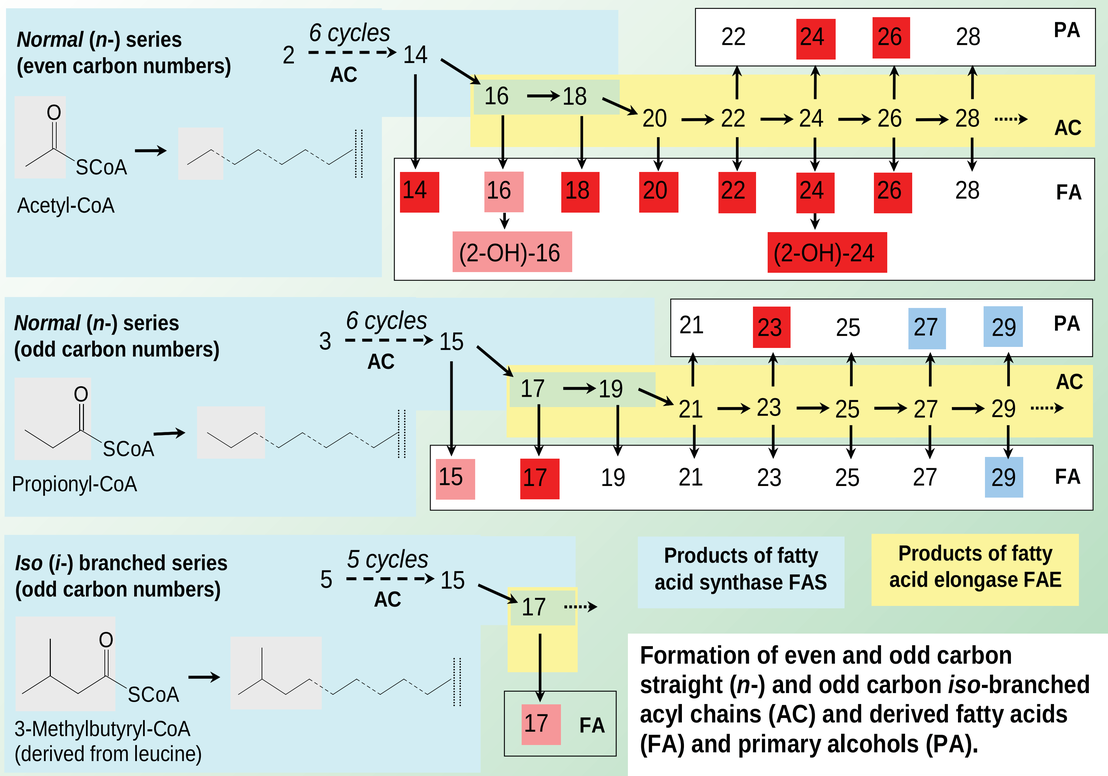
Analysis

Analysis of variance was performed on each metabolite separately, and using a cut-off of 2% false discovery rate 59 metabolites were identified as accumulating differently between the 4 groups. Hierarchical clustering, based on pairwise standard errors of difference between groups, partitioned these metabolites into clusters of similar significance patterns across the four groups.



Results

Three main groups of biosynthetically-related non-polar metabolites (composed of saturated fatty acids and fatty alcohols) were identified, differing predominantly in the length of their carbon chains and in the presence of chains with odd and even numbers of carbon atoms. The group containing shorter chain even carbon compounds also includes an odd carbon *iso*-branched fatty acid. The two graphical plots show the groups which include long chain odd-carbon fatty acids and alcohols, and the cluster which consists mainly of even-carbon longer chain fatty acid and alcohol homologues. The different patterns probably reflect differences in the specificity of the enzyme systems responsible for synthesis of long carbon chains, and also a shift in their activities or partitioning of precursors between the pathways in Phureja and Stenotomum in comparison with Tuberosum and Andigena. It believed that in many plant species parallel pathways probably exist for synthesis of odd and even carbon straight chain and branched chain components. Straight chain (*n*-) even carbon acids and alcohols are derived from sequential elongation of a C_2 starter unit derived from acetate, with C_2 units derived from malonate. Propionate is the primer for formation of *n*-odd carbon compounds, and leucine is the source of the C_5 precursor involved in formation of odd carbon *i*-branched compounds. Chain elongation up to C_{16}-C_{18} or C_{17}-C_{19} is catalysed by plastidial fatty acid synthase, FAS, in the synthesis denovo. Subsequent elongation to very long chain lengths is catalysed by an extra-plastidial fatty acid elongase FAE enzyme complex..



Conclusion

The value of this technique is reflected in the fact that the separation is not the same as seen with genetic analysis (Spooner et al, PNAS 102,14694) in that case the Phureja and the Andigena were more closely related. This reflects the difference in approaches with the genetic analysis looking at gene differences whereas metabolomics reports on the actual gene products themselves. It should be noted however that the environmental differences, i.e. growing such plants outwith their natural environment, will imcact upon their relative metabolites compositions.