Gene expression analysis of starch metabolism using mRNAseq and the potato genome sequence

Sønderkær, Mads; Kloosterman, Bjorn; Bachem, Christian; Nielsen, Kåre Lehmann

Publication date:
2010

Document Version
Accepted author manuscript, peer reviewed version

Gene expression analysis of starch metabolism using mRNAseq and the potato genome sequence

Mads Sønderkær, Bjorn Kloosterman**, Christian WB Bachem**, and Kåre Lehmann Nielsen**

**Department of Biotechnology, Chemistry and Environmental Engineering, Aalborg University*

***Wageningen UR Plant Breeding, Wageningen University and Research Centre*

mson@bio.aau.dk, Bjorn.Kloosterman@wur.nl, Christian.Bachem@wur.nl, kln@bio.aau.dk

Introduction

Crops such as potatoes that have storage organs (tubers) placed in the soil produce twice the amount of energy per area unit compared to cereals. This makes these kinds of crops well suited as a starting point for future crops for food and energy production. In order to develop a potato with higher starch yield than presently possible, detailed knowledge about starch metabolism is crucial. Accumulation of carbohydrates in the form of starch in potato tubers is the result of both anabolic and catabolic processes. These processes are highly redundant in terms of gene isoforms and multiple metabolic pathways. Synthesis of starch can take place by direct incorporation of glucose-1-phosphate into starch catalysed by starch phosphorylase [1] or via ADP-glucose catalysed by ADP-glucose pyrophosphorylase [2,3,4] and starch synthase [5], while starch breakdown can occur via phosphorylytic or hydrolytic reactions [6,7,8]. In potato, starch synthesis takes place not only in tubers but also in leaves in the form of transient starch during the day, which is consumed in the absence of photosynthesis during the night.

This poster will present the results of a transcriptome analysis based on the draft potato genome sequence v3. Samples from leafs, stolons and tubers from the two potato varieties; the diploid breeding hybrid RH (*S. tuberosum* x *tuberosum*) [9] and the doubled monoploid DM (*S. tuberosum* x *phureja*) [9],

Gene Model Curation

Table 1: Manual curation of 169 gene models participating in carbon metabolism.

Gene model types	No	Percentage
Gene models with correct transcript model	129	76.3 %
Gene models where the transcript having the longest CDS is correct	124	73.3 %
Gene models not validated due to low coverage and/or non unique matches	16	9.4 %
Gene models manually corrected based on mRNAseq	15	8.8 %
Gene models with more than 1 gene (gene fusion)	13	7.6 %
Correct gene model is annotated as multiple genes (split gene)	4	2.3 %

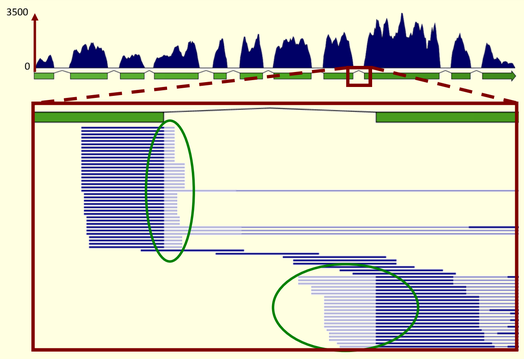

Figure 1: Example of manual curation of the exon-intron structure by mRNAseq. Illustration of an 11 exon sucrose synthase gene.
Top: The read coverage in the gene region matches the exon regions indicating exon-intron boundaries.
Bottom: mRNAseq reads partially matching (faded part) at the exon-intron boundary. The non-matching part of the read (green circles) matches the end of the sequence of the neighbouring exon, enabling the curation of the exon-intron boundary

The existing gene annotation (v 3.2) of the draft potato genome sequence (DM) has been automatically generated using *ab initio*, homology-based, and sequence assisted (EST and mRNAseq) methods resulting in 47,352 gene models. Of the subset containing 169 gene models used in this analysis, 76 % contained a transcript model where the exon-intron structure could be manually validated by mRNAseq data (Table 1) (Figure 1) indicating that the potato draft gene annotation is of equal or higher quality compared to the maize [10] and rice genome annotations [11]. Moreover, it seems that choosing the transcript model containing the longest CDS to represent each gene model having multiple transcripts predicted, only introduces a very small error. This approach will however cause true gene models to be deleted in cases where two gene models have been fused, which was observed in 7.6 % of the investigated cases.

Methods

Initially gene isoforms of the starch metabolism in the DM v3.2 genome sequence annotation were identified by BLASTP similarity search of know sequences, and by keyword search against the PGSC functional annotation database. The data set was expanded by: 1) Adding genes belonging to the same OrthoMCL cluster as a gene already included. 2) Adding genes having a significant BlastP score (e-value < 1e-20) to a gene already included.

The 38 DM mRNAseq libraries used for the transcriptomic assisted gene annotation of the DM v3 genome sequence and 17 RH libraries (818 mio. reads in total), were mapped to the DM v3 genome sequence. Mapping criteria was set to allow a maximum of 2 mismatches for short 36 bp reads, and 10 % mismatch plus a maximum distance of 10 kb between paired-end reads, the later chosen not to exclude exon-intron boundary spanning reads. The exon-intron structure of each gene isoform included in the data was curated by manually inspecting the reference assembly. In cases where a gene model had multiple transcripts, the longest transcript with validated exon-intron structure was chosen.

Expression levels for leaf, stolon and tuber samples in DM and RH were calculated by RNAseq analysis and reported as reads per kilobase of exon model per million mapped reads (RKPM) [12]. The normalized expression level for each gene was calculated relative to the highest expression observed for the gene. The absolute and relative expression levels was visualized as heat maps using the online visualization tool Prometra available at: https://prometra.cebitec.uni-bielefeld.de/cgi-bin/login.cgi

References

1. Fettke J., Albrecht T., Hejazi M., Mahlow S., Nakamura Y., and Steup M. (2010). Glucose 1-phosphate is efficiently taken up by potato (solanum tuberosum) tuber parenchyma cells and converted to reserve starch granules. *New Phytol.*, 185:663–675.
2. Müller-Röber B., Sonnewald U., and Willmitzer L. (1992). Inhibition of the ADP -glucose pyrophosphorylase in transgenic potatoes leads to sugar-storing tubers and influences tuber formation and expression of tuber storage protein genes. The EMBO Journal vol. 11(4):1229-1238.
3. Stark D.M., Timmerman K.P., Barry G.F., Preiss J., and Kishore G.M. (1992). Regulation of the Amount of Starch in Plant Tissues by ADP Glucose Pyrophosphorylase . Science 258:287-292.
4. Sweetlove L.J., Burrell M.M., and ap Rees T., (1996). Starch metabolism in tubers of transgenic potato (Solanum tuberosum) with increased ADPglucose pyrophosphorylase. Biochem. J. 320:493-498
5. Tetlow, I.J. (2006), Understanding storage starch biosynthesis in plants: A means to quality improvement, Canadian Journal of Botany, vol. 84, no. 8, pp. 1167-1185.
6. Kaplan, F., Sung, D.Y. & Guy, C.L. (2006), Roles of β-amylase and starch breakdown during temperatures stress, Physiologia Plantarum, vol. 126, no. 1, pp. 120-128.
7. Lloyd, J.R., Kossmann, J. & Ritte, G. (2005), Leaf starch degradation comes out of the shadows, Trends in plant science, vol. 10, no. 3, pp. 130-137.
8. Zeeman, S.C., Smith, S.M. & Smith, A.M. (2004), The breakdown of starch in leaves, New Phytologist, vol. 163, no. 2, pp. 247-261.
9. Visser, R.G.F., Bachem, C.W.B., de Boer, J.M., Bryan, G.J., Chakrabati, S.K., Feingold, S., Gromadka, R., van Ham, R.C.H.J., Huang, S., Jacobs, J.M.E., Kuznetsov, B., de Melo, P.E., Milbourne, D., Orjeda, G., Sagredo, B. & Tang, X. 2009, "Sequencing the Potato genome: Outline and first results to come from the Elucidation of the sequence of the world's third most important food crop", American Journal of Potato Research, vol. 86, no. 6, pp. 417-429.
10. Schnable, P.S., et. al 2009, The B73 maize genome: Complexity, diversity, and dynamics, Science, vol. 326, no. 5956, pp. 1112-1115.
11. Yu, J., et. al., (2002), A draft sequence of the rice genome (Oryza sativa L. ssp. indica), Science, vol. 296, no. 5565, pp. 79-92.
12. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. 2008, Mapping and quantifying mammalian transcriptomes by RNA-Seq, Nature Methods, vol. 5, no. 7, pp. 621-628.
13. Haake, V., Zrenner, R., Sonnewald, U. & Stitt, M. 1998, "A moderate decrease of plastid aldolase activity inhibits photosynthesis, alters the levels of sugars and starch, and inhibits growth of potato plants", Plant Journal, vol. 14, no. 2, pp. 147-157.

Starch Metabolism

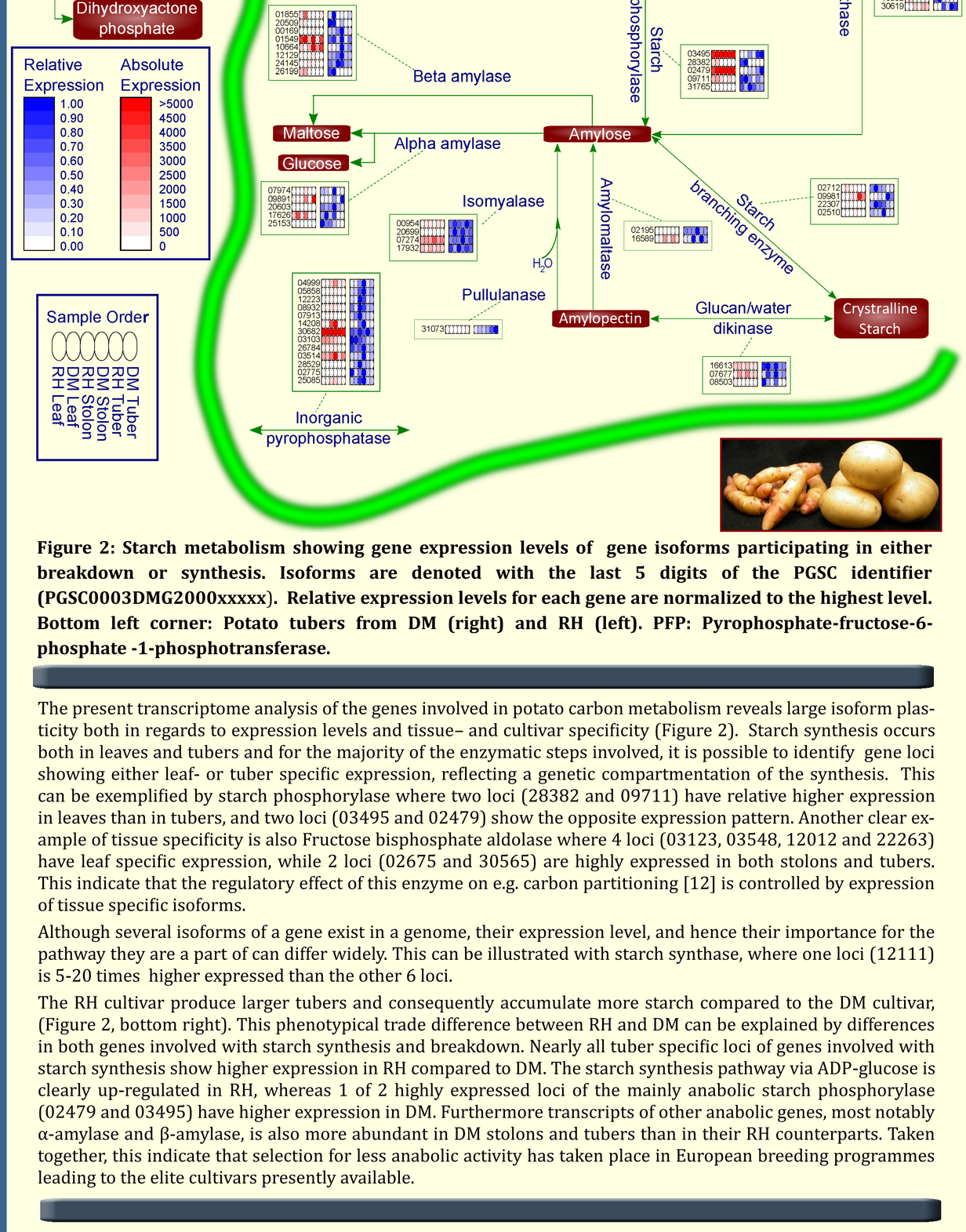

Figure 2: Starch metabolism showing gene expression levels of gene isoforms participating in either breakdown or synthesis. Isoforms are denoted with the last 5 digits of the PGSC identifier (PGSC0003DMG2000xxxxx). Relative expression levels for each gene are normalized to the highest level. Bottom left corner: Potato tubers from DM (right) and RH (left). PFP: Pyrophosphate-fructose-6-phosphate -1-phosphotransferase.

The present transcriptome analysis of the genes involved in potato carbon metabolism reveals large isoform plasticity both in regards to expression levels and tissue– and cultivar specificity (Figure 2). Starch synthesis occurs both in leaves and tubers and for the majority of the enzymatic steps involved, it is possible to identify gene loci showing either leaf- or tuber specific expression, reflecting a genetic compartmentation of the synthesis. This can be exemplified by starch phosphorylase where two loci (28382 and 09711) have relative higher expression in leaves than in tubers, and two loci (03495 and 02479) show the opposite expression pattern. Another clear example of tissue specificity is also Fructose bisphosphate aldolase where 4 loci (03123, 03548, 12012 and 22263) have leaf specific expression, while 2 loci (02675 and 30565) are highly expressed in both stolons and tubers. This indicate that the regulatory effect of this enzyme on e.g. carbon partitioning [12] is controlled by expression of tissue specific isoforms.

Although several isoforms of a gene exist in a genome, their expression level, and hence their importance for the pathway they are a part of can differ widely. This can be illustrated with starch synthase, where one loci (12111) is 5-20 times higher expressed than the other 6 loci.

The RH cultivar produce larger tubers and consequently accumulate more starch compared to the DM cultivar, (Figure 2, bottom right). This phenotypical trade difference between RH and DM can be explained by differences in both genes involved with starch synthesis and breakdown. Nearly all tuber specific loci of genes involved with starch synthesis show higher expression in RH compared to DM. The starch synthesis pathway via ADP-glucose is clearly up-regulated in RH, whereas 1 of 2 highly expressed loci of the mainly anabolic starch phosphorylase (02479 and 03495) have higher expression in DM. Furthermore transcripts of other anabolic genes, most notably α-amylase and β-amylase, is also more abundant in DM stolons and tubers than in their RH counterparts. Taken together, this indicate that selection for less anabolic activity has taken place in European breeding programmes leading to the elite cultivars presently available.

In conclusion, two important pieces of information can be deduced from this study.

1) The enzymes participating in leaf starch metabolism and tuber metabolisms is largely coded on separate genes. Hence it should be possible to alter metabolic activity in the tuber without severe effects on leaf starch metabolism.

2) Even though it at first seems a daunting task to manipulate the carbon flux in such a redundant pathway containing many gene isoforms in nearly all steps, most isoforms are likely to have little influence of the net reaction because of low expression and consequently low abundance of the enzymes they encode. Therefore by selecting the properly expressed isoforms, it should be possible to manipulate the carbon flux in the starch metabolism by altering the concentration of these key enzymes.

Acknowledgements

The authors gratefully acknowledge the Potato Genome Sequencing Consortium (PGSC) for providing the DM mRNA seq libraries.

This poster is supported by The Danish Council for Strategic Research, Programme Commission on Health, Food and Welfare, Grant 2101-07-0116